



Investigación educativa

5ª edición

James H. McMillan
Sally Schumacher



INVESTIGACIÓN EDUCATIVA 5.^a EDICIÓN
UNA INTRODUCCIÓN CONCEPTUAL

5.^a EDICIÓN

INVESTIGACIÓN EDUCATIVA

UNA INTRODUCCIÓN CONCEPTUAL

James H. McMillan
Sally Schumacher
Virginia Commonwealth University

Traducción

Joaquín Sánchez Baidés
Ingeniero informático

Revisión técnica y adaptación

Pilar Sánchez López
Universidad de Almería

y

Consuelo Clemente Pita
Editora escolar



Madrid • México • Santafé de Bogotá • Buenos Aires • Caracas • Lima
Montevideo • San Juan • San José • Santiago • São Paulo • White Plains

James H. McMillan-Sally Schumacher
Investigación educativa 5.ª edición
PEARSON EDUCACIÓN, S. A., Madrid, 2005

ISBN: 978-84-832-2687-2
Materia: Pedagogía 37.01

Formato: 17 x 24

Páginas: 664

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra sin contar con autorización de los titulares de la propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual (arts. 270 y sgts. del Código Penal).

DERECHOS RESERVADOS

© 2005 respecto de la primera edición en castellano por:
PEARSON EDUCACIÓN, S. A.
C/Ribera del Loira, 28
28042 Madrid (España)

PEARSON ADDISON WESLEY es un sello editorial autorizado de PEARSON EDUCACIÓN

James H. McMillan-Sally Schumacher

Investigación educativa 5.ª edición

ISBN: 84-205-4163-X

Depósito Legal: M-

Traducido de:

Research in Education. A conceptual introduction 5th edition

Copyright 2001 © Addison Wesley Longman

ISBN: 0-321-08087-4

Editor: Juan Luis Posadas

Técnico editorial: Elena Bazaco

Equipo de producción:

Director: José Antonio Clares

Técnico: Diego Marín

Diseño de cubierta: Departamento de Diseño de Pearson Educación, S. A.

Composición: DiScript Preimpresión, S. L.

Impreso por:

IMPRESO EN ESPAÑA - PRINTED IN SPAIN

Este libro ha sido impreso con papel y tintas ecológicos

Índice

PARTE I	
Principios fundamentales de la Investigación Educativa	1
Capítulo 1	
Introducción al campo de la Investigación Educativa	3
Capítulo 2	
Modalidades de investigación, técnicas de recogida de datos e informes de investigación	37
Capítulo 3	
Temas de investigación: enunciados, preguntas e hipótesis	87
PARTE II	
Diseños y métodos de investigación cuantitativa	127
Capítulo 4	
Introducción al diseño de investigación cuantitativa	129
Capítulo 5	
Estadística descriptiva	177
Capítulo 6	
Técnicas de recogida de datos cuantitativos	213
Capítulo 7	
Diseños de investigación no experimental y encuestas	267
Capítulo 8	
Diseños de investigación experimentales y de caso único	311
Capítulo 9	
Estadística inferencial	357
PARTE III	
Diseños y métodos de investigación cualitativa	397
Capítulo 10	
Introducción al diseño de investigación cualitativa	399
Capítulo 11	
Estrategias cualitativas	439
Capítulo 12	
Análisis de datos cualitativos	477

PARTE IV**Investigación analítica 521**

Capítulo 13

Análisis de conceptos e investigación histórica 523**PARTE V****Diseños y métodos de investigación política y evaluativa 555**

Capítulo 14

Investigación evaluativa y análisis político 557**Apéndice A****Directrices para proyectos de investigación 597****Apéndice B****Glosario 611****Apéndice C****Cálculos para estadísticos descriptivos e inferenciales seleccionados 635**

I

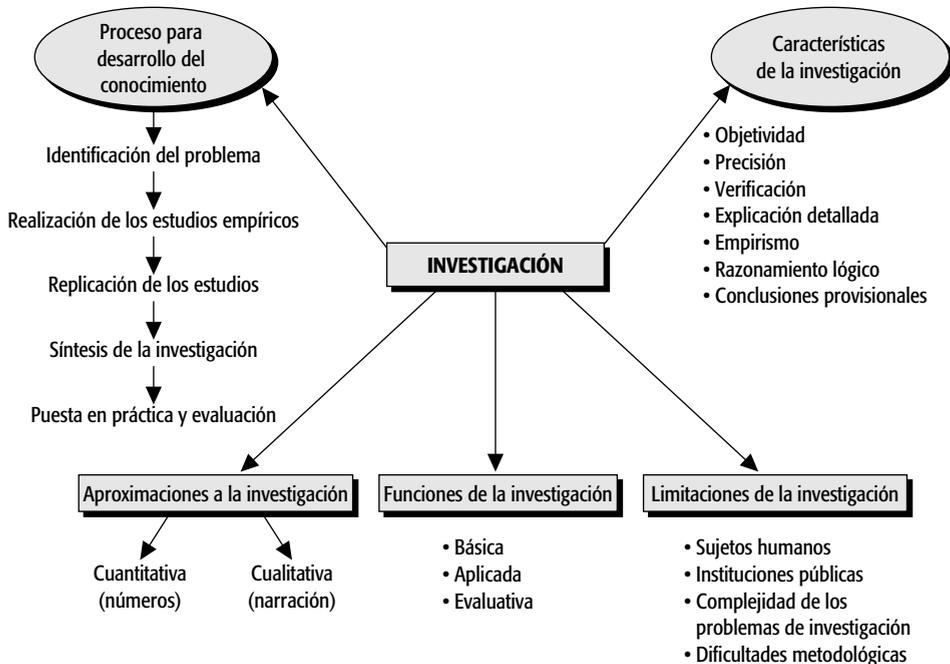
Principios fundamentales de la Investigación Educativa

- Qué es la investigación? Los educadores poco familiarizados con los métodos científicos plantean esta pregunta frecuentemente. También podrían preguntar: ¿por qué se considera que la investigación es más útil que la experiencia o el consejo de otros para la toma de decisiones? ¿Cómo influye la investigación en las prácticas educativas? ¿Qué tipo de estudios se hacen en educación? ¿Hay una manera sistemática de entender un artículo de investigación?

Los capítulos 1 y 2 responden a estas preguntas proporcionando una introducción al campo de la Investigación Educativa, una visión general de las modalidades de investigación y de las técnicas de recogida de datos y del formato de los artículos de investigación cuantitativa y cualitativa en las revistas especializadas. Esta introducción le familiarizará con la terminología básica y los conceptos fundamentales de la investigación. Todas las investigaciones comienzan con la formulación de un problema y, normalmente, implican una revisión bibliográfica. El capítulo 3 le ayudará a reconocer, enunciar y evaluar un problema de investigación. ¿Cómo se formula un problema para que sea útil en la planificación de un estudio? ¿Qué debería transmitir el enunciado de un problema a un lector? ¿Cómo se evalúan las formulaciones de los problemas?

Juntos, estos tres capítulos presentan los principios básicos de la investigación que un estudiante necesita comprender a la hora de realizar, leer y analizar diferentes tipos de investigación y metodologías. Las secciones siguientes abordarán con gran detalle los diseños y procedimientos para metodologías específicas.

Introducción al campo de la Investigación Educativa



PALABRAS CLAVE

replicación	empírico
síntesis de la investigación	datos
ciencia	razonamiento deductivo
teoría	razonamiento inductivo
estudio científico	investigación cuantitativa
método científico	investigación cualitativa
investigación	generalizabilidad
métodos de investigación	investigación básica
estudio sistemático	investigación aplicada
objetividad	investigación evaluativa
verificación	investigación-acción
explicación	análisis político

La Investigación Educativa ha afectado gradualmente a la mayoría de nuestras ideas sobre la educación y a las prácticas que utilizamos para lograr nuestros objetivos educativos. Aun así, muchos excelentes gestores y profesores saben poco sobre Investigación Educativa y asumen que la investigación no ha tenido efecto alguno en su actividad diaria. Los profesionales, también, suelen asumir que la investigación no es útil para el desarrollo de sus programas ni para la formulación de políticas.

Este capítulo introduce al lector en el ámbito de la Investigación Educativa describiendo el desarrollo del conocimiento para la mejora de la práctica educativa. La Investigación Educativa es un estudio científico y sistemático que utiliza aproximaciones cualitativas y cuantitativas. Los educadores y otros profesionales utilizan la investigación básica, aplicada y evaluativa para propósitos diferentes. Se describen algunas contribuciones y limitaciones de la Investigación Educativa. Y, lo más importante, presentamos el lenguaje y la lógica de la investigación empleados para leer y realizar estudios de investigación.

INVESTIGACIÓN EDUCATIVA EN EL SIGLO XX

Algunos titulares en el *Education Week*, un periódico de amplia difusión, ilustran la variedad de temas investigados recientemente.

- «*Study Links High School Courses Success*» (6/2/99, p. 5).
- «*Researchers Find Teacher Tests Short on Covering College Content*» (6/2/99, p. 5).

- «*More Teachers and Students Say Violence in Schools is Declining*» (6/2/99, p. 8).
- «*Study Highlights Benefits, Shortcomings of Magnet Programs*» (6/2/99, p. 7).
- «*Study: Effects of Child Care Linger in Early Grades*» (6/16/99, p. 3).
- «*Study Finds Mismatch Between California Standards and Assessments*» (6/23/99, p. 10).
- «*Finding Schools's Danger Zones*» (6/23/99, p. 38).
- «*Study: Voucher Parents Satisfied with Schools*» (6/23/99, p. 11).

La traducción de los títulos anteriores nos revela los asuntos, como la disminución de la violencia en los colegios o la satisfacción de los padres con la escuela, que ocupan el interés de la prensa acerca de los resultados de la Investigación Educativa.

Los últimos artículos de las revistas de investigación también reflejan la variedad de temas estudiados.

- ¿Por qué no van más alumnos afroamericanos a la escuela?
- Investigar a «mi gente», investigarme a mí mismo: fragmentos de una historia reflexiva.
- El profesor novel de ciencias: narrativas de convicciones y limitaciones.
- Principios de gobierno compartido: la experiencia interior.
- Inhibidores en la implementación de un enfoque de solución de problemas para la enseñanza de las ciencias: estudio de caso de un profesor en cambio.
- ¡Eh, esos zapatos no forman parte del uniforme!: niñas afroamericanas en una escuela de élite y la importancia de la vestimenta.
- Alternativas a la instrucción tradicional: el uso de juegos y simulaciones para aumentar el aprendizaje y la motivación de los alumnos.
- Prácticas programáticas de la escuela y satisfacción de los alumnos.
- Oferta y demanda de profesores pertenecientes a minorías en Texas: problemas y expectativas.
- Percepciones de los orientadores escolares acerca de la utilidad de los tests estandarizados, frecuencia de su uso y necesidades de entrenamiento para su evaluación.
- Impacto del tamaño de la clase en las estrategias de aprendizaje y el uso del tiempo en los cursos de matemáticas y ciencias en el instituto.
- Nunca volveremos atrás: educación de adultos y la lucha por la ciudadanía en el verano de la libertad de Misisipi.
- Maestros abrazando la magia: ¿cómo hacen uso de su conocimiento intuitivo los maestros eficaces?
- La investigación trasladada a prácticas para aumentar la implicación de los alumnos en actividades relacionadas con la transición.

- Juegos entre clases: los problemas de Estados Unidos con la clase, raza y género en un instituto y comunidad negros.
- ¿Quién dirigirá? Los 10 factores más destacados que influyen en los maestros a la hora de cambiar su puesto docente por otro en la administración.
- Estrés y resistencia: un estudio cualitativo de alumnos de 4º y 5º curso.
- Los efectos de las variaciones en el tamaño de la clase sobre las conductas: el caso de los profesores de matemáticas

Además, tres grupos nacionales han publicado informes en los que sugieren cuáles deben ser los temas prioritarios en los que la Investigación Educativa debe centrarse en el siglo XXI. Un conocido investigador educativo con más de 30 años de experiencia en el campo, L. B. Resnick, dijo que «hay más interés en el conocimiento que proviene de la Investigación Educativa del que pueda recordar en cualquier momento de mi carrera. Ahora todo el mundo se apunta al mismo carro ganador [preguntando] ¿Está basado en investigaciones? ¿Cuáles son las evidencias de las investigaciones?» (Viadero, 1999, p.1).

Importancia de la Investigación Educativa

¿Por qué la Investigación Educativa ha llegado a ser una fuente valiosa de información para una gran diversidad de grupos? Sugerimos algunas razones de este interés y reconocimiento de la Investigación Educativa.

Primero, *los educadores están constantemente intentando entender los procesos educativos y deben tomar decisiones profesionales*. Estas decisiones tienen efectos inmediatos y a largo plazo sobre otras personas: estudiantes, profesores, padres y, finalmente, nuestras comunidades y nuestra nación. ¿Cómo adquieren los educadores sus conocimientos para tomar decisiones? La mayoría de nosotros tendemos a confiar en muchas fuentes, incluyendo la experiencia personal, una opinión experta, la tradición, la intuición, el sentido común y nuestras creencias sobre lo que está bien y lo que está mal. Cada una de esas fuentes es legítima; sin embargo, en algunas situaciones, pueden ser inadecuadas como única referencia para tomar decisiones.

Segundo, *los estamentos políticos no educativos, como el estado y las autoridades e instituciones regionales, han ido introduciendo cada vez más cambios en la educación*. ¿Cómo adquieren los grupos políticos sus posiciones acerca de la educación y acceden a la información acerca de las escuelas y la enseñanza? La mayoría de los responsables de la toma de decisiones políticas prefiere disponer de información fundamentada en investigaciones relevantes para su uso político. Muchos de los responsables jurídicos estatales delegan en la Consejería Estatal de Educación la realización de estudios sobre las políticas educativas nacionales. Tanto las consejerías de educación regionales como las estatales subvencionan este tipo de estudios. Por esta razón, a los investigadores se les está pidiendo,

cada vez más, que trabajen en problemas complejos en entornos altamente politizados.

Tercero, *las personas interesadas, los profesionales y los grupos privados y fundaciones han incrementado sus actividades de investigación*. Las asociaciones educativas profesionales, los sindicatos de profesores, las asociaciones de padres de alumnos y las fundaciones como la *National Science Foundation* han realizado o han encargado estudios sobre temas de especial interés para la organización.

Cuarto, *las revisiones sobre investigaciones previas han interpretado la evidencia empírica acumulada*. Por ejemplo, estudios sobre repetición indican que tiene poco sentido, desde el punto de vista pedagógico, que un niño repita curso. Otras revisiones se han centrado en temas como el pensamiento en voz alta y la lectura comprensiva; los multimedia y la comprensión, el control y estilo del alumno; cómo se implican los padres en la educación de sus hijos; los parámetros del refuerzo positivo en la educación; la eficacia del profesor; los efectos de la escolarización separada por sexos o la coeducación en el desarrollo social, emocional y académico; o el estrés profesional, el síndrome de *burnout* y la salud del profesorado. Otras revisiones han servido para identificar áreas en las que se necesita investigar.

Quinto, *el fácil acceso a la Investigación Educativa*. La investigación acerca de la práctica educativa está siendo divulgada a través de las revistas científicas y profesionales, de los informes publicados por instituciones especializadas, libros, bases de datos de las bibliotecas, periódicos, televisión o Internet. Aunque la calidad de la investigación pueda variar dependiendo de la fuente, la Investigación Educativa es muy accesible.

Sexto, *muchos educadores, investigadores a tiempo parcial, consultan investigaciones y llevan a cabo estudios*. Este tipo de investigaciones ayudan a los educadores a planificar nuevos programas, mejorar su práctica educativa, evaluar el aprendizaje y asignar recursos a las necesidades cambiantes de sus propios entornos.

La información fiable es cada vez más necesaria en una sociedad técnica compleja. La investigación proporciona información válida y conocimientos precisos sobre la educación con el propósito de tomar decisiones informadas. En situaciones en que existen diversos puntos de vista, un estudio bien hecho aporta una perspectiva racional en los foros de discusión y en las mesas de elaboración de políticas. El proceso de investigación propone principios para ayudar a los educadores a tomar decisiones juiciosas.

Puesto que la investigación describe o mide fenómenos de manera sistemática, resulta una fuente de conocimiento mejor que las experiencias y opiniones particulares, que la tradición o la mera intuición. Algunos estudios son abstractos y proporcionan información general sobre prácticas y políticas educativas frecuentes. Este tipo de investigación influye en el modo en que uno piensa acerca

de la educación. Otros aportan información detallada sobre acciones concretas en un lugar específico, como una escuela, una clase o un programa. Este tipo de investigaciones puede utilizarse con carácter inmediato en la planificación, el desarrollo o la mejora de una práctica determinada.

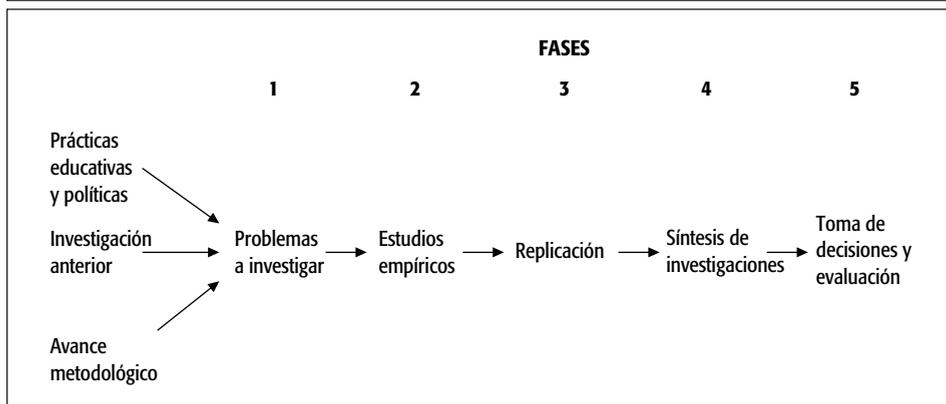
Desarrollo del conocimiento para mejorar la práctica educativa

El impacto de la investigación educativa en los colegios y en los políticos que tratan de mejorar la práctica educativa puede considerarse un proceso (Walberg, 1986). La figura 1.1 muestra las cinco fases del proceso de desarrollo del conocimiento educativo para mejorar la práctica. Estas fases son: (1) identificación de los problemas de investigación, (2) estudios empíricos, (3) replicación, (4) síntesis de la investigación y (5) puesta en práctica y evaluación.

Los problemas de investigación (fase 1) comienzan con una identificación de resultados a conseguir. Campos prácticos, como la educación, se ocupan de los resultados a conseguir, como el aprendizaje. Las preguntas y los problemas de investigación nacen de las siguientes fuentes: observación común, conocimiento práctico, controversias políticas, investigaciones anteriores y nuevas técnicas metodológicas aplicadas al estudio de la educación. Los investigadores llevan a cabo estudios empíricos (fase 2) y luego intentan **replicar**¹ los resultados obtenidos (fase 3) con diferentes sujetos y en escenarios y circunstancias distintos.

Aunque existen muchos intentos y son muy deseables, los estudios de réplica exacta son poco frecuentes en educación. Las **síntesis de investigaciones** y las

FIGURA 1.1: DESARROLLO DEL CONOCIMIENTO PARA MEJORAR LA PRÁCTICA EDUCATIVA



¹ Los términos en negrita son palabras clave y están definidos en el glosario.

revisiones evalúan sistemática y estadísticamente (fase 4) resumen de forma descriptiva estudios comparables. Un análisis de este tipo ayuda a organizar y dar sentido a los resultados globales de una investigación anterior. Así, más que unas cuantas réplicas exactas de una investigación original, es la suma de resultados de muchos estudios planificados cuidadosamente lo que constituye un conocimiento en el campo educativo basado en la investigación. Los encargados de la toma de decisiones y los responsables de ponerlas en práctica pueden aceptar razonablemente las implicaciones de los hallazgos de la investigación que son consistentes, «que son efectivos considerando los costes totales y no tienen efectos nocivos» (Walberg, 1986, p. 215). A pesar de ello, la evaluación local (fase 5) sigue siendo necesaria.

La investigación como búsqueda científica y sistemática

La investigación es una actividad relativamente nueva en la historia de la educación, de la misma manera que el concepto de escuela pública gratuita para todos los niños es relativamente reciente en la historia de la humanidad. En los primeros siglos, antes de que la lectura y la escritura fueran habituales, los individuos adquirían el conocimiento del mundo que les rodeaba de dos maneras. Una, a través de la experiencia personal y de la observación de las experiencias de otros. Después, comunicaban su conocimiento a la siguiente generación oralmente. Transmitían su sabiduría colectiva como una serie de descripciones detalladas de individuos y sucesos en situaciones y contextos concretos. Nadie esperaba que estas historias pudieran predecir sus futuras acciones, porque éstas dependían de las percepciones y sensaciones de una situación. La vida en aquellos tiempos era corta y arriesgada. Las historias, sin embargo, proporcionaban una comprensión, un bagaje cultural a partir del cual una persona podía extrapolar o aplicar experiencias ya conocidas a un área desconocida para llegar a conjeturas o imágenes válidas para el futuro.

El conocimiento se obtenía, también, de otra manera: a través de actividades humanas que se medían con números. Las primeras unidades de medida y escalas tenían una finalidad eminentemente práctica: medir de manera consistente y fiable la duración de un día, la distancia andada o la cantidad de espacio de un área. Asumiendo que las mismas unidades de medida simbolizan el mismo significado, no se necesitaba conocer los detalles descriptivos contextuales de la actividad medida. Un kilómetro era un kilómetro tanto sobre una superficie arenosa plana y lisa como sobre un terreno rocoso. Un pie equivalía a doce pulgadas tanto si se medía la altura de una persona, como la longitud de la piel de un animal o el interior de una cabaña. Si estas medidas eran fiables, se podían calcular segmentos de leyes naturales que hacían que los acontecimientos pudieran ser ordenados y predecibles.

En educación se necesitan diferentes clases de conocimiento y son muchos tipos de investigación los que, en el campo educativo, están orientados al desarrollo de este conocimiento. La Investigación Educativa se ha llamado búsqueda científica y búsqueda sistemática (ciencia social). Ambos puntos de vista están incluidos en nuestra definición de investigación y de las características de la Investigación Educativa.

El método científico

El fin último de la **ciencia** es la generación y verificación de una teoría. Una **teoría** predice y explica fenómenos naturales. En lugar de explicar todas y cada una de las conductas de los adolescentes por separado, por ejemplo, el científico busca explicaciones generales que subyacen a conductas diferentes. De forma similar, el científico, en lugar de intentar explicar los métodos de resolución de problemas matemáticos por los adolescentes, busca explicaciones generales a la resolución de problemas de todo tipo.

Un famoso investigador, Fred N. Kerlinger (1986), define *teoría* como un conjunto de constructos y proposiciones relacionadas que especifican relaciones entre variables para explicar y predecir fenómenos. Esta definición indica tres cosas acerca de una teoría. Primera, una teoría es un conjunto de proposiciones consistentes sobre constructos concretos. Segunda, una teoría pone de manifiesto las interrelaciones entre un conjunto de variables. Tercera, una teoría explica fenómenos. Al explicar qué variables se relacionan con qué otras y cómo lo hacen, un científico puede predecir unas variables a partir de otras. Si es posible predecir desde la variable A (test de ansiedad, por ejemplo) la variable B (test de rendimiento), entonces puede deducirse, por ejemplo, la posibilidad de «control» a partir de la intervención en, por ejemplo, las habilidades para hacer tests.

En el campo de las ciencias sociales, muchas investigaciones valiosas no están orientadas específicamente a la teoría. En cambio, esta investigación se concentra en los objetivos de menor alcance para encontrar relaciones específicas. Las relaciones más útiles y valiosas, sin embargo, son aquellas que son más generalizables, aquellas que están vinculadas a una teoría. Las relaciones específicas (como en el caso del test de ansiedad relacionado con el test de rendimiento), aunque son importantes e interesantes, son menos aplicables que si se encuentran en un «conjunto de variables interrelacionadas que son parte de una teoría» (Kerlinger, 1986, p.10). Las metas modestas, limitadas y específicas son muy válidas, pero la formulación y la verificación de teorías son mucho mejores porque son más generales y explicativas.

La generación y la verificación de una teoría son fundamentales para el método científico. Una teoría, para ser útil para el desarrollo del conocimiento científico, debe satisfacer ciertos criterios. Primero, una teoría debe proporcionar una

explicación simple de las relaciones relevantes observadas en un problema concreto. Segundo, una teoría debe ser consistente con las relaciones observadas y con el corpus de conocimientos ya existente. Un enunciado teórico es la descripción más probable y eficiente de la evidencia acumulada a través de investigaciones previas. Tercero, una teoría es considerada una explicación provisional y debe proporcionar medios para su verificación y revisión. Cuarto, una teoría debe estimular la investigación posterior en las áreas que lo necesiten. El análisis cuidadoso de una teoría y su apoyo empírico pueden proponer relaciones para estudiar en nuevas investigaciones.

Un científico valora la metodología empírica por «su forma de exponer a falsación, de todas las formas concebibles, el sistema que se quiere poner a prueba» (Popper, 1959, p. 42). El propósito no es promover un conocimiento falso, sino seleccionar cuál es el más exacto y fiable, sometiendo las teorías opuestas a la comprobación empírica. Es por esto que los estudios científicos se diferencian claramente de otros procedimientos en la búsqueda de un conocimiento válido y fiable.

La indagación científica es la búsqueda de conocimiento mediante el empleo de procedimientos válidos para la recogida, análisis e interpretación de datos. El término *científico* se refiere a una metodología y no es sinónimo de ciencia. Ciencia es un conjunto de conocimientos probados, mientras que «científico» se refiere a la forma en que ese conocimiento ha sido generado. El método científico es normalmente un proceso de investigación secuencial. Los pasos característicos del **método científico** son:

1. Planteamiento de un problema.
2. Formulación de las hipótesis que deben ser comprobadas.
3. Recogida y análisis de los datos.
4. Interpretación de los resultados y extracción de las conclusiones en relación con el problema.

El método científico es simplemente una metodología apta para el desarrollo de un conocimiento válido y fiable.

Definición de investigación

Definida en pocas palabras, una **investigación** es un proceso sistemático de recogida y de análisis lógico de información (datos) con un fin concreto. Esta definición es general porque existen muchos métodos disponibles para investigar un problema o una cuestión. La Investigación Educativa no se limita a las aproximaciones usadas en las ciencias físicas y naturales, aunque tampoco debería emplearse indiscriminadamente la palabra «investigación» para describir lo que es realmente una observación casual y una especulación. Los **métodos de**

investigación (a veces llamados «metodología») son las formas en que se recogen y analizan los datos. Estos métodos han sido desarrollados para adquirir conocimiento mediante procedimientos válidos y fiables. La recogida de datos puede hacerse con técnicas de medición, entrevistas y observaciones o documentos.

La metodología de una investigación es sistemática e intencional. Los procedimientos no son actividades casuales, sino que están planificados para obtener datos sobre un problema de una investigación particular. En un contexto amplio, la metodología se refiere a un diseño por medio del cual el investigador selecciona procedimientos de recogida y de análisis de los datos para investigar un problema específico. Es posible que un diseño no proporcione datos válidos o fiables sobre el problema, pero la elección deliberada de un diseño incrementa la probabilidad de que los datos aporten información sobre la cuestión investigada.

Investigación Educativa como búsqueda sistemática

Cuando hablamos de investigación en educación, nos referimos a un conjunto de métodos que comparten las características de la búsqueda sistemática. La búsqueda sistemática, al contrario de lo que se cree, «está dirigida y documentada de tal forma que el argumento puede ser examinado concienzudamente. El interés del informe no depende de la elocuencia del autor o de cualquier plausibilidad superficial» (Cronbach y Suppes, 1969, p. 15). «Sistemático» no se refiere a la ritualización y formas rígidas de investigación, ni a seguir necesariamente los procedimientos formales ya establecidos. Algunas de las búsquedas más brillantes son especulativas en sus fases iniciales, poniendo en práctica combinaciones de ideas y procedimientos o buscando incansablemente ideas. Cronbach y Suppes (1969, pp. 15-16) sugirieron:

Sea cual sea el carácter de la investigación, si es sistemática, el investigador ha debido prever cuáles son las preguntas habituales que son pertinentes. Establece un control en cada paso de la recogida de información y del razonamiento para evitar las fuentes de error con las que están relacionados. Si los errores no pueden ser eliminados, los tiene en cuenta para establecer el margen de error en sus conclusiones. De esta manera, el informe de un estudio sistemático tiene una estructura que presenta cuáles son las materias primas que conforman el problema y los procesos lógicos a los que han sido sometidas y reestructuradas para hacer creíble la conclusión.

L. S. Shulman (1981) señala varias razones por las que los investigadores educativos debaten sobre la adecuación de los métodos de investigación para estudiar los fenómenos. Primera, la Investigación Educativa exige seleccionar un conjunto específico de observaciones o hechos de entre infinitas posibilidades. Después

de todo, el número de instituciones y de actividades educativas son abundantes en la sociedad de los Estados Unidos. Al igual que en un juicio, los adversarios legales pueden mostrar su total desacuerdo sobre la relevancia de una prueba o la corrección del veredicto al que se llega tras razonar con esa evidencia; así es como se procede en un estudio sistemático en educación. A menudo existe una falta de consenso sobre los datos o sobre la evidencia y/o las cadenas de razonamiento.

Segunda, el estudio sistemático no se refiere sólo a la investigación sistemática sino también a las propias disciplinas. Materias como la psicología, la sociología, la historia, las ciencias políticas, la antropología y otras sirven de referencia para determinar cuáles son los principios de la investigación y los cánones de evidencia (datos) que debe utilizar el investigador. Los principios de la investigación son algo diferentes en cada una de las disciplinas.

Tercera, la principal razón por la que la metodología de investigación es un campo interesante en educación es que ésta en sí misma no es una ciencia o una disciplina, sino un área de estudio en el que los fenómenos, sucesos, personas, procesos e instituciones constituyen la materia prima para investigaciones de muchos tipos. Los enfoques y métodos de investigación propios de varias disciplinas pueden servir de ayuda para responder a las preguntas que se suscitan en el campo educativo y que son inherentes a él.

Las características de la Investigación Educativa

Las siguientes características son comunes a muchos tipos de la investigación que se realiza en educación: objetiva, precisa, verificable, explicativa, empírica, lógica y condicional. Todas ellas juntas describen la naturaleza de la investigación (tabla 1.1).

1. **Objetividad.** La objetividad es tanto un procedimiento como una característica. Para un abogado, objetividad significa imparcialidad, mente abierta, no subjetividad. Como procedimiento, la **objetividad** se refiere a la recogida de datos y a los procedimientos de análisis a partir de los cuales se puede obtener una interpretación razonable. La objetividad tiene que ver con la calidad de los datos conseguidos por los procedimientos de análisis que, o bien controlan sus sesgos o tienen en cuenta su subjetividad.

2. **Precisión.** La investigación usa un lenguaje técnico. El lenguaje técnico de la investigación se emplea no para confundir al lector sino para comunicar significados exactos. Expresiones como *validez* y *fiabilidad* en medición, *diseño de investigación*, *muestra aleatoria* y *significación estadística* hacen referencia a procedimientos técnicos. Otros términos como *comparación constante* y *reflexividad* se refieren a estrategias en la investigación cualitativa. Un lenguaje preciso describe acertadamente el estudio de tal manera que puede ser replicado y sus resultados pueden utilizarse de manera correcta.

TABLA 1.1: CARACTERÍSTICAS DE LA INVESTIGACIÓN EDUCATIVA

Características	Cuantitativa	Cualitativa
Objetividad	Descripción explícita de la recogida de datos y de los procedimientos de análisis.	Descripción explícita de la recogida de datos y de los procedimientos de análisis.
Precisión	Medida y estadísticos.	Descripción detallada de fenómenos.
Verificación	Resultados replicados por otros.	Extensión del entendimiento por otros.
Explicación detallada	Preferencia por las explicaciones menos complicadas.	Resumen de generalizaciones.
Empirismo	Datos numéricos.	Fuentes, evidencia.
Razonamiento lógico	Fundamentalmente deductiva.	Fundamentalmente inductiva.
Conclusiones provisionales	Enunciados de probabilidad estadística.	Resumen provisional de interpretaciones.

3. *Verificación.* Para desarrollar conocimiento es necesario diseñar y presentar un estudio que permita su **verificación**: los resultados podrán ser confirmados o revisados en una investigación posterior. Los resultados pueden verificarse de diferentes formas, dependiendo del propósito del estudio original. Si la investigación pretende poner a prueba una teoría, la comprobación con otros grupos o en otros escenarios puede confirmarla o revisarla.

Sin embargo, la mayoría de los estudios cualitativos proporcionan interpretaciones descriptivas sobre la situación seleccionada o el «caso». Estas interpretaciones quedan ampliadas, pero no replicadas, en investigaciones posteriores en situaciones similares. La investigación cualitativa no es verificada de la misma forma que la investigación cuantitativa.

La verificación o generalización se caracteriza, igualmente, por compartir los resultados de los estudios. Cualquier investigación tiene una dimensión social y sus informes se presentan a la comunidad profesional para su examen público. A través de este proceso de discusión progresiva, los investigadores desarrollan un corpus de conocimiento e identifican nuevas preguntas de investigación.

4. *Explicación detallada.* La investigación intenta explicar las relaciones entre los fenómenos y reducir la **explicación** a afirmaciones sencillas. La teoría «la frustración conduce a la agresión» es una explicación que predice y que puede ser puesta a prueba para su verificación. La afirmación de que «el aprendizaje del

profesor y el cambio curricular no pueden aislarse de la situación social en que se implementa el currículum» (Tobin y La Master, 1995) es una explicación que se puede investigar con más profundidad. La meta última de la investigación es, por lo tanto, reducir las realidades complejas a explicaciones sencillas².

5. **Empirismo.** La investigación se caracteriza por una postura y un enfoque claramente empírico. El término *empírico* tiene un significado técnico y otro lego. El significado lego de empírico es el que está guiado por la experiencia práctica, no por la investigación. Esta perspectiva pragmática afirma que si funciona, es correcto; independientemente de las causas, debe ser correcto porque funciona. Para el investigador, **empírico** significa guiado por la experiencia obtenida por métodos de investigación sistemáticos en vez de por opiniones o juicios de expertos. Una postura empirista requiere, generalmente, una suspensión temporal de la experiencia personal y de las opiniones. Los elementos críticos en la investigación son la evidencia y las interpretaciones lógicas basadas en ésta.

Para un investigador, evidencia es igual a **datos**, es decir, resultados obtenidos de la investigación de los que se extraen las interpretaciones o conclusiones. En líneas generales, los términos *datos*, *fuentes* y *evidencia* son empleados como sinónimos, para referirse a la información obtenida por métodos de investigación. Las puntuaciones de los tests y los listados de ordenador, las observaciones de campo y los registros de entrevistas, artilugios y documentos históricos son todos denominados datos.

6. **Razonamiento lógico.** Cualquier investigación requiere razonamiento lógico. El razonamiento es un proceso mental, que emplea reglas lógicas establecidas, por el que se llega de un enunciado general a una conclusión específica (deducción) o, a la inversa, de un enunciado concreto a una conclusión generalizable (inducción).

En el **razonamiento deductivo**, si las premisas son correctas, automáticamente la conclusión es correcta. No pueden darse conclusiones totalmente «nuevas» porque las premisas no se validan empíricamente. La lógica deductiva, sin embargo, puede identificar relaciones nuevas dentro del conocimiento ya existente.

En el **razonamiento inductivo**, un investigador alcanza una conclusión mediante la observación de casos concretos (individuos, situaciones, sucesos) y formulando generalizaciones. De esta manera, las conclusiones se restringen a los casos particulares observados. Ningún sistema de razonamiento lógico es totalmente satisfactorio, pero cuando ambos se integran en un proceso de investigación, el estudio resulta más eficaz.

7. **Conclusiones provisionales.** Una concepción errónea acerca de la investigación es creer que los resultados son absolutos y las conclusiones verdaderas más allá de toda duda. No es verdad. Como apuntó un destacado investigador educativo, «La ciencia conductista y la investigación no ofrecen certezas. (¡Ni las

² Los investigadores cualitativos discuten si las explicaciones son proposiciones, asertos, generalizaciones naturalistas/concisas o conclusiones. Ver Stake (1995), Lincoln y Guba (1985) y Le Compte y Preissle (1993).

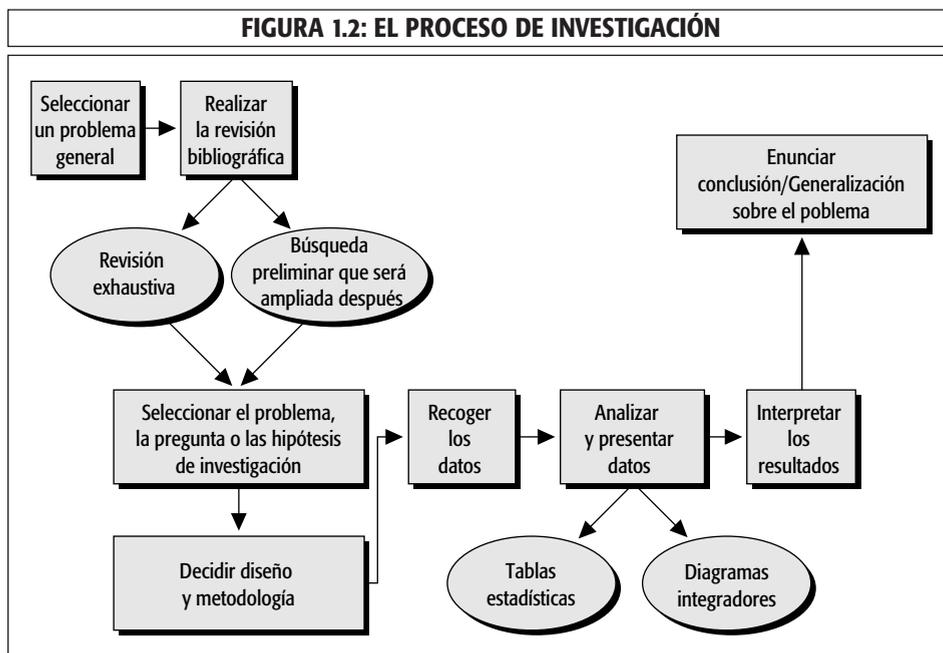
ciencias naturales!). Ni siquiera una certeza relativa. Todo lo que ofrecen es conocimiento probabilístico. Si se hace A, entonces probablemente ocurrirá B» (Kerlinger, 1979, p. 28). Una forma de definir la investigación puede ser decir que es un método de reducción de la incertidumbre. Las ciencias sociales tienen más incertidumbre que las ciencias físicas.

Arrojar conclusiones provisionales resulta fundamental para la investigación. Todas las disciplinas científicas y aplicadas contienen interpretaciones restringidas. Los enunciados de ambos tipos de investigación, cuantitativa y cualitativa, poseen unas conclusiones provisionales implícitas o explícitas. De esta manera, los investigadores escriben a menudo que sus resultados «parecen indicar» o «sugieren que».

El proceso de investigación

El proceso de investigación incluye, normalmente, varias fases. Estas fases no son secuenciales ni constituyen un proceso ordenado paso a paso. La investigación es más un proceso interactivo entre el investigador y la lógica del problema, el diseño y las interpretaciones. A continuación se presenta un resumen del proceso con algunas variaciones señaladas (figura 1.2).

1. **Selección de un problema general.** El problema define el área educativa en la que se centrará la investigación, por ejemplo, instrucción, administración, educación de adultos o educación especial.



2. **Revisión de la bibliografía sobre el problema.** La bibliografía más importante está formada por la investigación anterior y la teoría, pero también puede resultar útil otra bibliografía. En algunos estudios, se hace una revisión exhaustiva de la literatura antes de la recogida de datos. En otros, esta revisión es provisional y preliminar a la recogida de los datos y luego, una vez recogidos, se amplía.

3. **Formulación del problema, objetivos o hipótesis específicos de la investigación.** Requiere que el investigador decida si es más apropiada para la investigación la modalidad cuantitativa o cualitativa. Si se selecciona una metodología cualitativa, los problemas o las preguntas de investigación sirven de guía inicial y se irán especificando a medida que la investigación avance.

4. **Determinación del diseño y la metodología.** El investigador decide de qué sujetos se van a obtener los datos, cómo se van a seleccionar dichos sujetos y cómo se van a recoger los datos.

5. **Recogida de los datos.** Deben resolverse, además, aspectos éticos y legales con respecto a la recogida de los datos y su análisis.

6. **Análisis de los datos y presentación de los resultados.** Normalmente se emplean las representaciones visuales gráficas como tablas estadísticas o diagramas integradores.

7. **Interpretación de los resultados y conclusiones o presentación de las generalizaciones respecto al problema.** Hay que tomar decisiones sobre el formato apropiado del informe, en función de la finalidad de la investigación y de los destinatarios o lectores. El proceso de investigación puede ser relativamente corto o puede llevar varios años.

Cualquier investigación es, fundamentalmente, una búsqueda reflexiva, en la que cada decisión adoptada por el investigador se describe explícitamente, justificando racionalmente su elección. Es un proceso intelectual apasionante en el que se utilizan diferentes habilidades en las diferentes fases.

ENFOQUES DE LA INVESTIGACIÓN CUANTITATIVA Y CUALITATIVA

Los términos «cuantitativo» y «cualitativo» se utilizan frecuentemente para identificar diferentes modalidades de estudio o aproximación a la investigación. Cuando se emplean estos términos, se hace en dos niveles de discurso. En un primer nivel, cuantitativo y cualitativo³ se refieren a la distinción sobre la naturaleza del conocimiento: cómo entiende el investigador el mundo y el objetivo

³ La investigación cualitativa también suele denominarse «interpretativa», en referencia al estudio de los significados inmediatos y particulares de las acciones sociales de las personas (Erickson, 1986).

último de la investigación. En otro nivel, los términos se refieren a los métodos de investigación (cómo se recogen y analizan los datos) y al tipo de generalizaciones y representaciones que se derivan de ellos.

En educación se utilizan ambos enfoques, el cualitativo y el cuantitativo. La forma más evidente para el lector de distinguir entre una investigación cuantitativa y cualitativa es la forma de presentación de los datos. La **investigación cuantitativa** presenta resultados estadísticos en forma de números; la **investigación cualitativa** presenta los datos como una narración. Las diferencias, sin embargo, van más allá de la forma en que se presentan los datos (Denzin y Lincoln, eds., 1994)⁴. Los puristas sugieren que los métodos de investigación cuantitativa y cualitativa se basan en diferentes concepciones del mundo, en el objetivo de la investigación, en los métodos de investigación, en estudios característicos, en el papel del investigador y en la importancia del contexto en el estudio (tabla 1.2).

TABLA 1.2: APROXIMACIONES CUANTITATIVA Y CUALITATIVA A LA INVESTIGACIÓN

Orientación	Cuantitativa	Cualitativa
Concepción del mundo	Una realidad única, ej.: medida mediante una prueba.	Realidades múltiples, ej.: entrevistas al director, profesores y alumnos sobre una situación social.
Objetivo de la investigación	Establecer relaciones entre variables medidas.	Comprensión de una situación social desde la perspectiva de los participantes.
Procesos y métodos de investigación	Los procedimientos (pasos secuenciales) son establecidos antes de que comience el estudio.	Estrategias flexibles y cambiantes; el diseño emerge a medida que se recogen los datos.
Estudio prototípico (ejemplo más claro)	Diseño experimental para reducir el error y los sesgos.	Etnografía que emplea una «subjetividad sistematizada».
Papel del investigador	Desvinculado por el empleo de una prueba.	Personas preparadas se integran en la situación social.
Importancia del contexto	Meta: generalizaciones universales libres de contexto.	Meta: generalizaciones detalladas vinculadas al contexto.

⁴ Ver N. K. Denzin y Y. S. Lincoln, eds. (1994), *Handbook of qualitative research*.

1. **Concepciones del mundo.** La investigación cuantitativa, normalmente, está basada en alguna forma de «positivismo lógico», el cual asume que hay hechos sociales estables, con una realidad única, separados de los sentimientos y de las opiniones de los individuos. Por el contrario, la investigación cualitativa está más basada en el «construccionismo» que asume realidades múltiples construidas socialmente a través de las percepciones o puntos de vista individuales y colectivos diferentes de la misma situación.

2. **Objetivo de la investigación.** La investigación cuantitativa busca establecer relaciones y explicar las *causas* de los cambios en los hechos sociales medidos. La investigación cualitativa está más preocupada en la *comprensión* de los fenómenos sociales desde la perspectiva de los participantes. Esto ocurre a través de la participación, hasta cierto punto, del investigador en la vida de los sujetos durante la investigación.

3. **Métodos y procesos de investigación.** En los estudios cuantitativos se recurre a un conjunto de procedimientos y pasos establecidos que guían al investigador. En los estudios cualitativos, hay una gran flexibilidad tanto en las estrategias como en el proceso de investigación. Habitualmente, el investigador cualitativo utiliza un *diseño emergente* y revisa las decisiones sobre las estrategias de recogida de datos durante el proceso; por el contrario, los investigadores cuantitativos eligen los métodos como parte de un diseño *preestablecido* antes de la recogida de datos.

4. **Estudios característicos o prototípicos.** El investigador cuantitativo emplea diseños experimentales o correlacionales para reducir el error, el sesgo y las posibles variables extrañas. El estudio cualitativo prototípico de los hechos en curso es de índole *etnográfica*, lo que ayuda a los lectores a entender las perspectivas múltiples de la situación según las personas estudiadas. Mientras la investigación cuantitativa busca controlar el sesgo a través del diseño, la investigación cualitativa busca tener en cuenta la subjetividad en el análisis e interpretación de los datos.

5. **Papel del investigador.** El investigador cuantitativo ideal *permanece al margen* del estudio para evitar los sesgos. Los investigadores cualitativos llegan a estar *inmersos* en la situación y en los fenómenos que están estudiando. Por ejemplo, los investigadores cualitativos asumen papeles sociales interactivos en los que registran observaciones y entrevistas con los participantes en una amplia variedad de situaciones. Los investigadores cualitativos ponen de relieve la importancia de la recogida de datos por una *persona* preparada experta frente a un *instrumento*. La investigación cualitativa destaca por su «subjetividad sistematizada» (Erickson, 1973) y «reflexividad» (Mason, 1996), es decir, por el autoexamen crítico del papel del investigador durante todo el proceso⁵.

⁵ Algunas veces llamado «inserción del yo» en el texto para diferenciarlo de la recogida de datos de los participantes. Ver M. Fine (1998).

6. *Importancia del contexto en el estudio.* La mayoría de la investigación cuantitativa intenta establecer *generalizaciones universales libres de contexto*. El investigador cualitativo cree que las acciones de los seres humanos están fuertemente influenciadas por los escenarios en los que tienen lugar. Además, como afirma Wilson (1977) «aquéllos que trabajan dentro de esta tradición [cualitativa] afirman que el científico social no puede entender la conducta humana sin comprender el marco dentro del que los sujetos viven sus pensamientos, sentimientos y acciones» (p. 249). Este marco o contexto es descrito durante la recolección y el análisis de datos por el investigador. La investigación cualitativa desarrolla *generalizaciones ligadas al contexto*.

Muchas de estas distinciones entre investigación cuantitativa y cualitativa no son absolutas cuando se lleva a cabo un estudio o se lee uno ya acabado. Los investigadores experimentados pueden, y de hecho lo hacen, combinar ambos métodos de investigación, cuantitativo y cualitativo, en un mismo estudio para investigar un problema concreto (capítulo 14). Sin embargo, combinar ambos enfoques en un estudio es más difícil de lo que pudiera parecer (Tashakkori y Teddlie, 1998). No obstante, las distinciones son útiles en una introducción a la investigación para describir y entender los métodos de estudio, meta última de este libro. Las diferencias entre los enfoques se tratan en el capítulo 2 y se explican con detalle en la parte II, Métodos y diseños de investigación cuantitativa, y en la parte III, Métodos y diseños de investigación cualitativa.

LAS FUNCIONES DE LA INVESTIGACIÓN

La investigación hace avanzar el conocimiento y mejora la práctica. Esta simple afirmación plantea múltiples interrogantes. ¿Qué significa hacer avanzar el conocimiento, es decir, desarrollar un corpus de conocimiento científico? ¿Cómo puede el conocimiento científico, que a menudo se caracteriza por ser poco aplicable, mejorar la práctica educativa? Estas preguntas no son nuevas para investigadores, lectores o usuarios de la investigación. Quizá, hoy en día son más relevantes y apremiantes dado el aumento del uso de los resultados de la investigación. Algunos conceptos erróneos sobre las funciones de la investigación han llevado a más de una crítica injustificada hacia los esfuerzos de la investigación y el conocimiento científico.

Considere algunos de los usos posibles de la investigación en situaciones educativas. Suponga que en un estudio de valoración de necesidades descubre que en una escuela muchos alumnos leen por debajo del nivel correspondiente a su curso y muy pocos consiguen puntuaciones altas en una prueba estandarizada. ¿Debería esta escuela emplear sus limitados recursos en un programa para lectores avanzados o, por el contrario, dedicarlos a los lectores más atrasados?

Imagínese que un administrador lee que los niños de bajo rendimiento al recibir «recompensas masivas» puntúan más alto en matemáticas que el grupo de control. ¿Significa esto que el administrador debería pedir al profesor que use un programa de refuerzo positivo con cada uno de los niños que rinde menos de lo esperado? Suponga que un tutor lee un estudio científico que verifica la teoría de que la frustración conduce a la agresión. ¿Puede esta teoría explicar por qué un estudiante de bachillerato que perdió un trabajo a tiempo parcial y sacó malas notas en los exámenes de acceso a la universidad, no fuera elegido por el equipo deportivo por sus deficientes resultados académicos y fuera enviado al despacho del director por pelearse? ¿Esta teoría, verificada por un solo estudio, le dice al tutor qué hacer en una situación así?

Una forma útil de entender cómo la investigación hace avanzar el conocimiento y mejora la práctica es analizar las funciones o el uso que se hace de diferentes tipos de investigación. Clasificamos las funciones de investigación como *básica*, *aplicada* y *evaluativa*. La investigación básica es llamada a veces *investigación pura o fundamental*. Las diferencias entre investigación básica, aplicada y evaluativa radican esencialmente en las funciones, especialmente en el grado en que facilitan la toma de decisiones⁶.

Las diferencias entre investigación básica, aplicada y evaluativa se reflejan en los temas de investigación, los objetivos, el nivel de discurso y la generalizabilidad de las explicaciones y en el empleo que se pretende dar al estudio (tabla 1.3). **Generalizabilidad** es la amplitud con que los hallazgos de un estudio pueden ser utilizados como conocimiento de otras poblaciones y situaciones, es decir, su capacidad de predicción. La comprensión de las funciones de la investigación básica, aplicada y evaluativa ayuda al diseño, la redacción y la utilización de la investigación. Muchos estudios son diseñados y considerados adecuados para un tipo de investigación y menos adecuados para otro. Raramente un mismo estudio se considera igualmente adecuado para los tres tipos de investigación.

Primero, necesitamos definir *campo*, *disciplina* y *ciencia*. Campo, en este ámbito, es un área de investigación, conocimiento y práctica que es más que una mera disciplina académica, como la medicina o la educación. Disciplina se refiere a un método de organización del conocimiento académico. Las disciplinas se clasifican, normalmente, como ciencias físicas, ciencias sociales y del comportamiento. Aunque los investigadores y científicos debaten si una disciplina concreta es una ciencia, en general, están de acuerdo en que es el tema de investigación el que distingue las ciencias físicas, sociales y de la conducta.

⁶ Los autores reconocen que las diferencias entre investigación básica, aplicada y evaluativa, tal y como se presentan aquí, probablemente estén demasiado simplificadas y exageradas dado su mero carácter ilustrativo.

TABLA 1.3: FUNCIONES DE LOS TIPOS DE INVESTIGACIÓN

	Básica	Aplicada	Evaluativa
Tema de la investigación	1. Ciencias físicas, sociales y del comportamiento.	1. Campo aplicado: medicina, ingeniería, educación.	1. Prácticas en situaciones concretas.
Propósito	1. Probar teorías, leyes científicas, principios básicos. 2. Establecer relaciones empíricas entre fenómenos y generalizaciones analíticas.	1. Probar la utilidad de las teorías científicas en un campo determinado. 2. Establecer relaciones empíricas y generalizaciones analíticas en un campo determinado.	1. Evaluar el mérito de una práctica específica. 2. Evaluar el coste de una práctica específica.
Nivel de discurso/ Generalizabilidad	1. Abstracto, general.	1. General, asociado a un campo determinado.	1. Concreto, específico a una práctica concreta. 2. Aplicado a una práctica específica en una situación determinada.
Uso perseguido	1. Aumentar las leyes y principios básicos del conocimiento científico. 2. Mejorar los procedimientos y metodologías.	1. Aumentar el conocimiento científico en un campo determinado. 2. Avanzar en la investigación y metodología en un campo determinado.	1. Aumentar el conocimiento científico de una práctica específica. 2. Avanzar en la investigación y metodología de una práctica específica. 3. Ayudar en la toma de decisiones en una situación determinada.

Investigación básica

Investigación básica es la investigación realizada para probar una teoría con escasa o ninguna intención de aplicar sus resultados a problemas prácticos. La investigación básica, que se preocupa exclusivamente por saber, explicar y predecir fenómenos sociales y naturales, empieza con una teoría, un principio básico o una generalización. Una teoría puede tener o no soporte empírico. Cuando una teoría tiene un considerable apoyo empírico se denomina ley científica. Una ley científica, como la gravedad, es generalizable, es decir, explica muchos casos individuales.

En la investigación básica, la abstracción es necesaria para cualquier teoría. Por ejemplo, un enunciado abstracto como «el refuerzo favorece el mantenimiento del aprendizaje» es valioso porque abarca muchas manifestaciones de refuerzo y de mantenimiento. La abstracción, como parte del poder de la ciencia, está siempre alejada de las preocupaciones diarias y de la calidez de las relaciones humanas.

La investigación básica no está diseñada para resolver problemas sociales. El científico se preocupa por el desarrollo del conocimiento y no se le exige que explique las implicaciones prácticas de un estudio. Ninguno de estos objetivos puede ser alcanzado con un único estudio. Sin embargo, la investigación básica acaba influyendo en las prácticas sociales una vez transcurrido un considerable periodo de tiempo, ya que un nuevo conocimiento supone un reto para el conjunto de dogmas y creencias. La investigación básica afecta indirectamente a la forma de pensar y de percibir de la gente, pero no necesariamente les lleva a actuar.

Por lo tanto, el objetivo de la investigación básica es primero incrementar nuestro conocimiento de los principios básicos y las leyes científicas, y segundo hacer progresar los procedimientos y las metodologías científicas. Los científicos pueden poner a prueba teorías organizativas en contextos educativos para desarrollar teorías de administración educativa, o bien pueden comprobar teorías sobre el aprendizaje, el desarrollo cognitivo, la creatividad y otros programas instructivos.

Investigación aplicada

La **investigación aplicada** se centra en un campo de práctica habitual y se preocupa por el desarrollo y la aplicación del conocimiento obtenido en la investigación sobre dicha práctica. La medicina, la ingeniería, el trabajo social o la educación son campos aplicados que se valen del conocimiento científico, pero no son ciencias en sí mismas. La investigación aplicada (al contrario que la básica) alcanza un conocimiento relevante para dar solución (generalizable) a un problema *general*. En otras palabras, los estudios aplicados se centran en los problemas de investigación habituales en un campo determinado.

La investigación aplicada comprueba la utilidad de las teorías científicas y determina las relaciones empíricas y analíticas dentro de un campo determinado. Un científico, por ejemplo, puede investigar la inteligencia humana en muchas situaciones diferentes, incluida la escolar. Un investigador educativo puede determinar la relación entre inteligencia y rendimiento o analizar la política federal sobre el uso educativo de ciertos textos, pero los resultados de estos estudios se formulan con carácter general y no constituyen recomendaciones específicas para la actuación inmediata.

La investigación aplicada, como la investigación básica, es abstracta y general, y emplea un lenguaje común dentro de un campo. Muchos estudios educativos sobre rendimiento, enseñanza, aprendizaje, niños excepcionales y asuntos parecidos utilizan términos que tienen un significado especial dentro de la profesión. La generalizabilidad de la investigación aplicada se limita normalmente al campo descrito. La investigación educativa, por tanto, se centra más en el conocimiento de teorías y de prácticas *educativas* que en un conocimiento *universal*.

La investigación aplicada incrementa el conocimiento obtenido mediante la investigación en un campo concreto. Los efectos de la investigación aplicada se notan indirectamente a largo plazo. Cuando se publica y discute una serie de estudios durante un considerable periodo de tiempo, el conocimiento que proporciona influye en la manera en que los profesionales enfocan y perciben un problema común. Los estudios aplicados estimulan la investigación posterior, sugieren nuevas teorías acerca de la práctica y pueden alentar el desarrollo metodológico. Sirva de ejemplo el desarrollo de estadísticas más apropiadas para el estudio de los fenómenos relacionados con el rendimiento escolar.

Investigación evaluativa

La investigación evaluativa se ocupa de una práctica concreta en una situación o situaciones determinadas. La práctica puede ser un programa, un producto o un proceso, pero la situación es crucial. La **investigación evaluativa** juzga el *mérito* y el *coste* de una práctica concreta en función de los valores que priman en la situación. La evaluación determina si la práctica funciona, es decir, ¿consigue lo que pretende en esa situación? Además, también analiza si los costes que suponen su desarrollo, aplicación y extensión compensan. Los costes pueden ser los materiales, el espacio, el personal implicado, el estado de ánimo del profesor y/o el apoyo de la comunidad.

La evaluación formal puede ser llevada a cabo por un investigador en un campo determinado o por un participante en la situación. La investigación evaluativa requiere, normalmente, un entrenamiento previo en distintas metodologías y disciplinas así como en habilidades comunicativas e interpersonales. Muchas

evaluaciones sistemáticas simultanean datos cuantitativos y cualitativos obtenidos en estudios realizados en distintas fases de la implantación de la práctica.

Los resultados de un estudio evaluativo se comunican en un lenguaje concreto, propio de la práctica, y compartido por los participantes. Dado que la evaluación ayuda a la toma inmediata de decisiones, sus resultados tienen normalmente una generalizabilidad limitada. Ambos tipos de evaluación, básica y evaluativa, pueden estudiar prácticas comunes a muchas situaciones educativas, pero se centran fundamentalmente en los intereses y problemas relacionados con la práctica en una situación determinada.

La investigación evaluativa puede incrementar nuestro conocimiento sobre una práctica específica, además de estimular el desarrollo de la metodología para estudiarla y la investigación posterior. Una serie de estudios de evolución de la práctica específica en lugares diversos, como las aulas del título IV (programa *Head Start*)⁷ o los procesos de cambio en las organizaciones pueden ampliar el conocimiento existente en el campo al que se aplica.

Variantes de la investigación evaluativa. Las variantes de la investigación evaluativa incluyen la investigación-acción y los estudios políticos. La **investigación-acción** implica a profesores que usan métodos de investigación para estudiar los problemas en el aula. El profesor dirige el estudio o desempeña un papel importante en el proceso de investigación. Puesto que el objetivo es dar solución a un problema concreto en un lugar específico, no es esencial un control riguroso de la investigación. Una variación muy reciente es la investigación-acción colaborativa (Oja y Smulyan, 1989; Stinger, 1996). En ella los participantes realizan el estudio con la ayuda de un asesor. La investigación-acción colaborativa normalmente se centra tanto en los procesos como en los resultados de un cambio de estrategia, como un programa para el desarrollo del personal.

Los **análisis políticos** evalúan los planes de la Administración para proporcionar recomendaciones prácticas a los responsables políticos. La investigación política se centra en las políticas escolares presentes y pasadas de gobiernos locales, estatales y federales. El análisis político (1) investiga la elaboración de planes, decidiendo especialmente qué temas educativos tratar; (2) examina la implementación de programas para llevar a cabo políticas; y (3) determina su eficacia y/o eficiencia. Muchos estudios investigan varias alternativas de planes dentro de las realidades económicas y políticas existentes.

Resumiendo, las investigaciones básicas, aplicadas y evaluativas se diferencian en primer lugar, porque tienen distintos propósitos, temas de investigación y usos perseguidos. Ni la investigación básica ni la aplicada están diseñadas

⁷ *Head Start* es un programa existente en EE.UU. a nivel federal, de desarrollo y asistencia social para niños de bajos recursos económicos en edad preescolar y para sus familias, desarrollo infantil de salud y de asociaciones comunitarias.

para obtener soluciones inmediatas a un problema específico en una situación determinada. Sin embargo, ambos tipos de investigación pueden influir indirectamente, pasado un periodo de tiempo considerable, en la forma en que los científicos, investigadores y participantes afrontan y perciben sus respectivos problemas. La investigación evaluativa juzga el mérito y el coste de una práctica particular en una situación determinada. A diferencia de la investigación básica y de la aplicada, los estudios evaluativos proporcionan información para su uso inmediato cuando las directrices políticas y prácticas se desarrollan, implementan e institucionalizan⁸. Los tres tipos de investigación son válidos para el desarrollo del conocimiento.

Recordemos también que la *función* (propósito) y la *calidad* son dos dimensiones diferentes de la investigación. Los investigadores utilizan los mismos tipos de diseños y métodos para los diferentes tipos de investigación. El criterio para determinar la calidad de un estudio está relacionado con el diseño y los procedimientos. Puede haber investigaciones básicas mal diseñadas y excelentes estudios aplicados. La investigación-acción y el análisis político a pequeña escala pueden estar bien diseñados; al mismo tiempo, estudios evaluativos a gran escala pueden proporcionar información cuestionable a causa de dificultades de procedimiento.

LA EDUCACIÓN COMO CAMPO DE INVESTIGACIÓN Y PRÁCTICA

La educación es un área de investigación interdisciplinar que proporciona descripciones, explicaciones, predicciones y evaluaciones de las prácticas educativas. La práctica educativa se centra en la enseñanza y el aprendizaje, e incluye factores que influyen en la enseñanza, como el desarrollo del currículum, las innovaciones, la administración, el desarrollo del docente y las políticas educativas. El conocimiento basado en la investigación refleja esta dualidad de la educación como un campo de estudio y de práctica.

Campo de investigación interdisciplinar

La educación, como campo de investigación interdisciplinar, ha tomado prestados conceptos y teorías de la psicología, la sociología, la antropología, las ciencias políticas, la economía y de otras disciplinas. Teorías que emplean conceptos como rol, norma, estatus, poder, autoridad, efectividad de costes, autoestima, desarrollo humano, cultura y otros semejantes han sido probadas en las prácticas

⁸ Ver Cousins y Leithwood (1986) con respecto a los factores que intervienen en la utilización de la evaluación desde una revisión de los estudios empíricos sobre esta cuestión.

educativas y han dado origen a nuevos conceptos de carácter educativo. Cuando se aplicó el análisis económico a la educación, por ejemplo, el concepto de ganancia fue sustituido por el de coste-beneficio, lo que significa que no todos los resultados educativos pueden ser traducidos a términos monetarios.

La investigación educativa también ha tomado prestadas metodologías desarrolladas originalmente por las ciencias sociales. La psicología, sobre todo la medición, ha dominado la investigación educativa tradicionalmente y continúa ejerciendo una fuerte influencia. Los investigadores también valoran otras metodologías como los estudios sociológicos, la observación participante de la antropología, la investigación histórica y el análisis político. Algunos de estos enfoques se aplican directamente a la educación, mientras que otros se modifican a la hora de investigar un problema educativo. Sin embargo, tales modificaciones rara vez llegan a quebrantar el significado que tenían en aquellas disciplinas de las que provienen.

El uso de diferentes conceptos y metodologías procedentes de distintas disciplinas enriquece y amplía el conocimiento basado en la investigación educativa. Los investigadores pueden estudiar casi cualquier tema con diferentes métodos. Un asunto relacionado con las ciencias de la educación puede, por ejemplo, ser investigado mediante un estudio de las necesidades curriculares de un entorno concreto; el estudio comparativo del rendimiento de alumnos con métodos de enseñanza diferentes; la observación de la interacción dentro de un aula; el estudio histórico de libros de texto de ciencias, o el análisis etnográfico de los procesos instructivos en clases de ciencias específicas. Cada una de estas aproximaciones amplía nuestro conocimiento sobre la enseñanza de las ciencias.

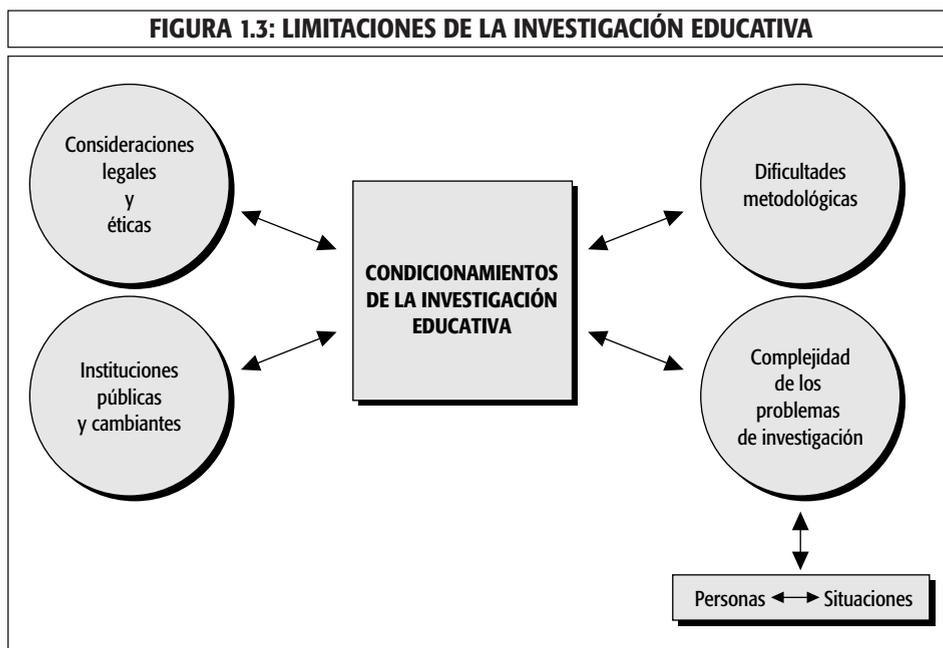
En un campo interdisciplinario, todas las metodologías de investigación son valoradas por su utilidad potencial para el desarrollo del conocimiento. Algunos investigadores pueden pensar que una metodología es mejor que otra, pero esto resulta engañoso porque cualquier metodología tiene sus propias limitaciones. Diferentes metodologías y distintos diseños proporcionan tipos diferentes de conocimiento sobre las prácticas educativas y, en algunos estudios, proporcionan métodos complementarios para investigar un tema (Jaeger, 1988).

Limitaciones de la Investigación Educativa

Desde la década de los treinta (Daniel, 1999), los investigadores se han preocupado por la calidad de la investigación educativa. A menudo, la educación es comparada con la medicina y el papel de los estudios científicos sobre la política de salud y la práctica sanitaria. Sin embargo, los investigadores del campo educativo sostienen que ambos campos y sus limitaciones no son análogos. Junto a numerosas restricciones institucionales y metodológicas, el coste es «la mayor preocupación en un campo (educación) que, en general, está considerado como infrafinanciado» (Viadero, 1999b, p. 34; ver Biddle, 1996).

El conocimiento adquirido a través de la investigación es limitado por la naturaleza tanto de la práctica educativa como de la investigación. La Investigación Educativa está condicionada por consideraciones éticas y legales ya que investiga con seres humanos, por la propia naturaleza pública de la educación, por la complejidad de las prácticas educativas y por las limitaciones metodológicas. Cada una de estas cuatro restricciones influye en última instancia en nuestro conocimiento sobre la educación (figura 1.3).

1. **Sujetos humanos.** La investigación educativa se centra, esencialmente, en los seres humanos. El investigador es responsable desde el punto de vista ético de salvaguardar los derechos y el bienestar de los sujetos mientras realiza un estudio. Debe protegerlos de las incomodidades físicas y psicológicas, perjuicios y situaciones peligrosas. Muchos estudios requieren el consentimiento previo de los sujetos, padres o instituciones y existen leyes para proteger la confidencialidad de los datos y la privacidad del individuo⁹. Durante años, los investigadores se han atenido de manera informal a estos principios que, a menudo, imponen limitaciones sobre los tipos de estudios que pueden realizarse legalmente en educación. El malestar físico y psicológico de los sujetos puede afectar a la duración de los periodos de administración de pruebas, a la réplica de



⁹ Las leyes, aprobadas en 1974, son *The Family Education Rights y Privacy Act*, *The National Research Act*, y *Privacy Act*.

los estudios, a los tipos de tratamientos y, en última instancia, a las cuestiones investigadas.

2. **Institución pública.** La educación es un servicio público influenciado por el ambiente externo. Aunque muchas prácticas parecen no experimentar cambios drásticos, las propias instituciones varían. Una escuela de primaria puede reemplazar por completo su profesorado y su población estudiantil cada seis años. La comunidad que rodea a una escuela puede modificarse drásticamente cuando una industria o un grupo étnico se establecen en la zona. Las leyes y las normativas legales han cambiado la estructura de los colegios y han incluido o eliminado programas. Dado que la finalidad de los colegios es la educación, y no la investigación, los estudios deben interferir lo menos posible en los procesos educativos. Los estudios longitudinales y los estudios de réplica de instituciones y personas en continuo cambio son difíciles de llevar a cabo. A menudo se dice que realmente no podemos conocer los efectos de la escolarización porque pueden ocurrir años más tarde fuera del escenario educativo.

La naturaleza pública de la educación también influye en los tipos de temas que se investigan. Muchos estudios hacen que los sujetos y otros grupos sean conscientes del tema de investigación. Algunos temas pueden ser demasiado controvertidos para una comunidad conservadora o producir desavenencias en los órganos directivos. Algunos estudios no son llevados a cabo debido a las posibles reacciones posteriores que pueden ser perjudiciales para el mantenimiento de una organización educativa.

3. **Complejidad de los problemas investigados.** Una tercera restricción en la investigación educativa es la complejidad de los problemas investigados. Las personas involucradas (alumnos, profesores, administradores, padres, comunidad educativa) son seres humanos complejos y seleccionan de forma activa los elementos a los que responden. Además, individuos distintos procesan ideas de manera diferente. La investigación educativa ha mostrado la complejidad de las disparidades individuales. El investigador, pues, trata simultáneamente con muchas variables, a menudo ambiguas, en un mismo estudio.

Además, muchos científicos sociales creen que los individuos no pueden ser estudiados ignorando el contexto de la vida real. La conducta está determinada tanto por características individuales como situacionales, y estudiar a los individuos sin considerar éstas últimas es incompleto. De esta forma, los investigadores educativos deben tener en cuenta no sólo las diferencias individuales entre las personas, sino que también deben considerar un gran número de elementos situacionales.

4. **Dificultades metodológicas.** Una cuarta limitación se refiere a las dificultades metodológicas. La investigación educativa mide características humanas complejas, como el pensamiento y las habilidades de resolución de problemas. La medición del rendimiento, de la inteligencia, de los estilos de liderazgo, de la interacción grupal o de las habilidades lectoras implica definiciones conceptuales

y cuestiones de validez. Algunas investigaciones sólo han sido posibles cuando se han elaborado pruebas válidas y fiables. Otras dificultades son el uso de una metodología inapropiada y las prácticas de divulgación de los resultados (Daniel, 1999; Thompson, 1996).

La investigación cualitativa también posee dificultades metodológicas, especialmente aquéllas inherentes al empleo de estrategias multi-método, al papel reflexivo de la investigación y a la explicitación de las técnicas de análisis de datos. A menudo, se ha criticado a la investigación cualitativa desde un punto de vista convencional por la falta de resultados fiables y generalizables, pero el estudio de casos proporciona generalizaciones ligadas al contexto para investigaciones futuras (Peshkin, 1993).

Relevancia de la Investigación Educativa

A pesar de estas dificultades, la investigación ha conseguido logros importantes en el campo del conocimiento. Algunas áreas de la Investigación Educativa tienen una larga tradición en la resolución de problemas metodológicos. Las metodologías empleadas más recientemente están ampliando el conocimiento acumulado. La importancia de la Investigación Educativa se manifiesta en parte por el incremento del número de revistas de Investigación Educativa y de estudios publicados en revistas profesionales. Cada vez con más frecuencia, nuestra sociedad y los responsables de la educación utilizan la investigación para ayudar en la toma de decisiones y el diseño de políticas educativas.

En una sociedad tecnológica, es necesaria una información válida para tomar decisiones educativas. Es más probable que la información basada en la investigación defina el problema cuidadosamente y refleje la complejidad de los procesos educativos. Si se planifica mejorar la educación, el primer paso es tener una información y conocimiento válidos.

La investigación influye en las instituciones y en los roles de los educadores de muchas maneras. Los educadores, con frecuencia, no son conscientes de que utilizan los resultados de una investigación educativa o de que han realizado alguna parte del proceso de investigación a lo largo de sus vidas profesionales. Las actividades enumeradas más abajo ilustran los usos más comunes de la Investigación Educativa en las instituciones y programas educativos. No es una lista exhaustiva.

1. Adquirir una perspectiva nueva o diferente sobre la educación o sobre un proceso educativo que genere ideas sobre cómo abordar un problema práctico.
2. Utilizar los resultados de la investigación para ayudar en la toma de decisiones y en la elaboración de políticas educativas, así como para justificar decisiones adoptadas entre varias alternativas.

3. Reconocer que muchas de las «innovaciones» muy en boga son similares a las que se aplicaron hace veinte o cuarenta años.
4. Ser capaz de seleccionar, a partir de una revisión bibliográfica, aquellos programas que tengan una mayor probabilidad de producir los efectos deseados.
5. Determinar la siguiente fase del desarrollo de un programa basándose en un informe de evaluación.
6. Ser capaz de leer críticamente estudios de investigación.
7. Distinguir las propuestas educativas realizadas por las autoridades y las declaraciones de organizaciones de aquéllas basadas en investigaciones.
8. Identificar necesidades instructivas o educativas de manera sistemática en vez de hacerlo de forma intuitiva.
9. Poner a prueba un supuesto o una hipótesis deducida de una teoría o de una propuesta que se traduzca en un conocimiento más fiable y en implicaciones para la práctica educativa.
10. Interpretar adecuadamente los resultados de un test estandarizado aplicado a los alumnos de un programa.
11. Administrar correctamente pruebas estandarizadas, grupales o individuales, para garantizar la validez y fiabilidad de los datos.
12. Responder a las peticiones para participar en un proyecto de investigación.

Con un conocimiento conceptual de la investigación, los interesados pueden leer, dirigir estudios a pequeña escala y desempeñar un papel activo en proyectos de investigación. Con un entrenamiento adicional y la experiencia, pueden diseñar y dirigir estudios que permitan el avance del conocimiento y la mejora de la práctica educativa.

RESUMEN

Este capítulo ha abordado el desarrollo del conocimiento educativo, la investigación científica y sistemática, las características de la investigación en la educación, las funciones de la investigación básica, aplicada y evaluativa y ha descrito la educación como un campo de estudio interdisciplinar y aplicado. Las ideas principales del capítulo se resumen a continuación.

1. La investigación proporciona una fuente para el conocimiento y la toma de decisiones mejor que la experiencia personal, las opiniones, la tradición o la simple intuición.
2. El proceso de desarrollo de un conocimiento educativo fiable implica la identificación de los problemas de investigación, los estudios empíricos,

los estudios de réplica y las síntesis de investigación, así como su aplicación y evaluación.

3. La ciencia aspira a la generación y verificación de teorías: un conjunto de proposiciones interrelacionadas que especifica relaciones entre variables para explicar y predecir un fenómeno.
4. Una teoría debe proporcionar una explicación sencilla de las relaciones observadas, tener en cuenta la evidencia ya existente, verificarla y estimular la investigación posterior.
5. La Investigación Educativa es un estudio sistemático en el que diferentes disciplinas proporcionan principios de investigación diferentes mediante los que recoger y analizar los datos.
6. La investigación es un proceso sistemático de recogida y análisis lógico de datos con algún objetivo.
7. La investigación se caracteriza por ser objetiva, precisa, verificable, explicativa, empírica, lógica y provisional.
8. Tanto la investigación cuantitativa como la cualitativa, utilizadas ambas en el campo educativo, difieren en sus concepciones sobre el mundo, en el propósito de la investigación, en los métodos y procesos, en el tipo de estudios, en el papel del investigador y en la importancia concedida al contexto.
9. La investigación básica pone a prueba teorías y explica las relaciones empíricas y analíticas en las ciencias físicas, sociales y de la conducta.
10. La investigación aplicada pone a prueba la utilidad de las teorías científicas en un campo aplicado e investiga las relaciones y las generalizaciones analíticas comunes a una profesión determinada.
11. Ni la investigación básica ni la aplicada están diseñadas para proporcionar soluciones directas a un problema específico en una situación concreta. Ambas pueden influir, indirectamente, transcurrido un cierto tiempo en el modo en que los científicos, investigadores y profesionales analizan y perciben sus respectivos problemas.
12. La investigación evaluativa juzga el mérito y el coste de una práctica concreta en una situación determinada con una o más escalas de valores. Entre las variantes de este tipo de investigación se incluyen la investigación-acción y el análisis político.
13. La calidad de una investigación depende de los métodos y del diseño, con independencia de que sea básica, aplicada o evaluativa.
14. La educación es un campo interdisciplinario de investigación, es decir, los investigadores toman prestados conceptos, teorías y metodologías de otras disciplinas y los aplican a su campo.
15. El conocimiento educativo está condicionado por factores éticos y legales, por la naturaleza pública de la educación, por la complejidad de las prácticas educativas y por limitaciones metodológicas.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. El conocimiento basado en la recogida sistemática de datos y su análisis es una...
 - a. opinión.
 - b. investigación.
 - c. experiencia personal.
 - d. tradición.
2. Una teoría, generada y verificada por la investigación, es útil porque...
 - a. proporciona una explicación sencilla de las relaciones observadas.
 - b. da cuenta de la evidencia anterior.
 - c. es suficientemente generalizable como para abarcar muchos casos individuales y manifestaciones.
 - d. todo lo anterior es correcto.
3. La investigación...
 - a. es el uso de métodos sistemáticos y explícitos y de razonamiento lógico que pueden analizarse críticamente.
 - b. se limita a los procesos de medición y análisis estadísticos.
 - c. depende de las características personales del investigador.
 - d. se caracteriza por ser incuestionable o absoluta, no admite dudas.
4. La investigación cuantitativa y cualitativa son similares en...
 - a. el papel del investigador.
 - b. el tipo de estudios.
 - c. las concepciones del mundo.
 - d. el énfasis de los datos empíricos.

Las afirmaciones de los ítems 5 a 7 hacen referencia a diferentes tipos de investigación. Empareje cada una de las afirmaciones con el tipo de investigación.

- a. Investigación básica.
 - b. Investigación aplicada.
 - c. Investigación evaluativa.
 - d. Investigación-acción.
5. Pone a prueba teorías y explica relaciones en las ciencias físicas, sociales y de la conducta.
 6. Evalúa el mérito o valor de una práctica concreta en una situación determinada.
 7. Pone a prueba la utilidad de las teorías científicas en un campo determinado.
 8. El conocimiento abstracto más general se deriva de la...
 - a. investigación evaluativa de un programa.

- b. investigación aplicada en un campo determinado.
 - c. investigación básica en las ciencias físicas, sociales y de la conducta.
 - d. investigación educativa.
9. Una explicación –es decir, un enunciado de relaciones entre casos particulares en un amplio conjunto de situaciones– *no* es una...
- a. teoría.
 - b. ley científica.
 - c. generalización analítica.
 - d. opinión de un experto.
10. La investigación educativa *no* está condicionada por...
- a. las restricciones éticas y legales.
 - b. la dependencia de una metodología.
 - c. las dificultades metodológicas.
 - d. la complejidad de las actividades de las personas.

Problemas aplicados

A continuación se presentan algunos ejemplos de resultados de investigaciones. ¿Para qué pueden servir esos resultados? Elija una de las siguientes respuestas y explíquela. Puede haber más de una respuesta correcta.

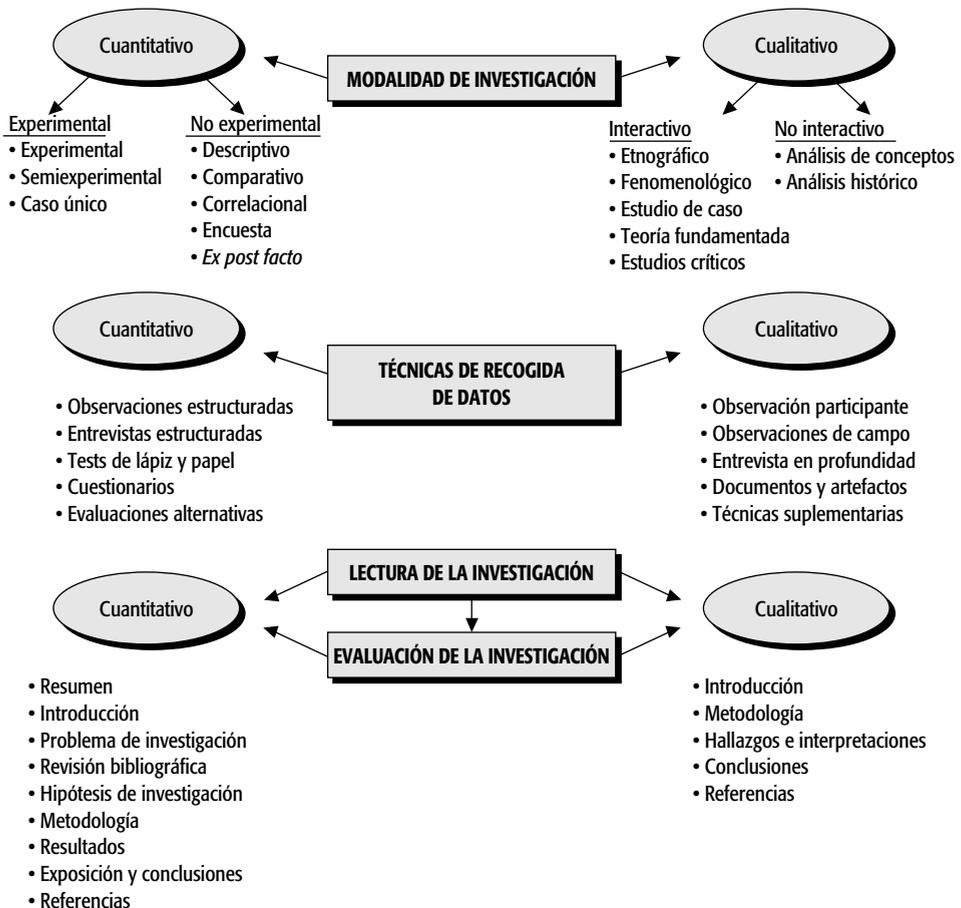
- a. Para influir en el modo en que el lector analiza o percibe un problema.
 - b. Para favorecer la toma de decisiones que lleve a la acción.
 - c. Para generar nuevos temas o problemas de investigación.
1. Un profesor lee las conclusiones de una investigación que afirma que los niños de hogares desestructurados manifiestan en el colegio un mayor número de conductas disruptivas que los que proceden de hogares estructurados.
2. Un estudio informa que una prueba de comprensión lectora para los cursos de 1º a 4º se ha validado en los alumnos de 1º y 2º pero no en los de 3º y 4º¹⁰.
3. Un investigador en historia de la educación observa que un estudio bien conocido sobre la organización de las escuelas públicas entre los años 1900-1950 se interrumpió por una sentencia de la Corte Suprema de 1954 sobre «separado pero igual».
4. Un director de un colegio comprueba, a través de los resultados de una encuesta, que los padres de sus alumnos no entienden ni el nuevo boletín de notas ni la escala de evaluación.
5. Un responsable de planificación curricular diseña un estudio piloto para ayudar a los profesores de educación de adultos a enseñar estrategias de

¹⁰ Según el sistema estadounidense.

lectura. Los resultados de una muestra representativa de los profesores sugieren que se revise el estudio de modo que incluya un fundamento teórico, una especificación clara del tipo de estudiante que debe beneficiarse de las estrategias y técnicas alternativas para responder a las dificultades que se les presenten a los alumnos.

6. La investigación previa indica que el sistema escolar ha sido una organización estructurada con una autoridad jerarquizada. Un administrador escolar recuerda que varios inspectores y directores han observado numerosos rasgos de conductas autónomas por parte de profesores y equipos directivos de centros escolares aunque no haya estudios empíricos que avalen estas apreciaciones.

Modalidades de investigación, técnicas de recogida de datos e informes de investigación



PALABRAS CLAVE

modalidad de investigación	estudio de caso
diseño de investigación	teoría fundamentada
experimental	estudios críticos
asignación aleatoria	investigación analítica
semiexperimental	análisis de concepto
caso único	análisis histórico
no experimental	observaciones estructuradas
descriptivo	entrevista estructurada
comparativo	cuestionarios
investigación correlacional	tests de lápiz y papel
correlación	evaluaciones alternativas
investigación mediante encuesta	observación participante
<i>ex post facto</i>	observación de campo
investigación cualitativa	entrevista en profundidad
etnografía	documentos
estudio fenomenológico	técnicas suplementarias

Este capítulo completa nuestra visión general del campo de la investigación educativa. Sus objetivos son presentar la metodología relacionada con la forma en que la investigación es diseñada y los datos son recogidos, y familiarizarle con la organización de informes de investigación publicados. Abarcaremos todas las modalidades y técnicas de investigación con más detalle en capítulos posteriores. Nuestra experiencia al enseñar investigación es que es mejor llegar a estar familiarizado con estos términos y conceptos lo antes posible. Al estar revisados en el contexto de estudios reales y explicados en detalle, usted ganará una comprensión más completa y un mejor recuerdo.

MODALIDADES DE INVESTIGACIÓN

Hemos examinado cómo la investigación se puede entender como una búsqueda científica y sistemática que puede ser, en primer lugar, cuantitativa o cualitativa, y que en la investigación se pueden distinguir diversas categorías: básica, aplicada o evaluativa. Otra forma de considerar la investigación esta basada en la *modalidad de investigación* del estudio. Una **modalidad de investigación** es una colección de prácticas eclécticas de indagación basada en un conjunto general de suposiciones, e implica preferencias metodológicas, opiniones filosóficas e

ideológicas, cuestiones de investigación y resultados con viabilidad. Otros autores pueden denominar a estas modalidades de investigación *tradiciones de investigación*.

Una modalidad de investigación informa sobre el diseño de exploración más apropiado. Un **diseño de investigación** describe los procedimientos para guiar el estudio, incluyendo cuándo, de quién y bajo qué condiciones serán obtenidos los datos. En otras palabras, el diseño indica cómo se prepara la investigación, qué le pasa a los sujetos y qué métodos de recogida de datos se utilizan.

El propósito de un diseño de investigación es proporcionar, dentro de una modalidad de investigación apropiada, las respuestas más válidas y exactas a las preguntas que se plantean. Dado que existen muchos tipos de preguntas de investigación y muchos tipos de diseños, es importante coordinar el diseño con la pregunta. El diseño de investigación es una parte muy importante de una investigación, en tanto que con cada diseño se relacionan determinadas limitaciones y reservas a la hora de interpretar los resultados y, también, porque el diseño de investigación determina cómo deberían ser analizados los datos.

Para ayudarle a entender mejor las modalidades de investigación, primero las hemos clasificado en dos grandes tipos: *cuantitativa* y *cualitativa*. Están incluidas dos categorías principales de las modalidades de investigación cuantitativas (experimental y no experimental) y dos categorías de las cualitativas (interactiva y no interactiva). Presentaremos los diseños más habituales en cada una de estas categorías y los discutiremos con más detalle en los capítulos posteriores.

La tabla 2.1 presenta las principales modalidades de investigación. Cada tipo comunica un aspecto diferente de la naturaleza del estudio. Se debería apuntar que estas características son independientes de la clasificación de la investigación como científica o sistemática o como básica, aplicada o evaluativa; esto es, por ejemplo, que la investigación básica puede ser experimental o no experimental y la investigación aplicada puede ser de caso único o correlacional.

MODALIDADES DE INVESTIGACIÓN CUANTITATIVA

Las modalidades de investigación cuantitativa fueron inicialmente desarrolladas para la investigación en agricultura y en las ciencias puras. Estos campos de estudio adoptaban una filosofía positivista del conocimiento que destacaba la objetividad y la cuantificación de los fenómenos. Como resultado, los diseños de investigación maximizan la objetividad con el empleo de números, de estadísticas, de estructura y de control experimental.

TABLA 2.1: MODALIDADES DE INVESTIGACIÓN

Cuantitativa		Cualitativa	
Experimental	No experimental	Interactiva	No interactiva
Experimental	Descriptiva	Etnográfica	Análisis de conceptos
Semiexperimental	Comparativa	Fenomenológica	Análisis histórico
Caso único	Correlacional Encuesta ^a <i>Ex post facto</i>	Estudio de caso Teoría fundamentada Estudios críticos	

^aLas encuestas aquí están clasificadas como un diseño de investigación. Las encuestas también están consideradas como un tipo de técnica de recogida de datos.

Modalidades de investigación experimental

En una modalidad experimental, el investigador manipula lo que los sujetos experimentarán. En otras palabras, el investigador tiene algún control sobre lo que les pasará a los sujetos mediante la imposición u ocultación sistemáticas de condiciones específicas. El investigador, entonces, establece comparaciones entre los sujetos que han soportado o que no han soportado las condiciones impuestas o entre sujetos que han experimentado condiciones diferentes. Las modalidades de investigación experimental tienen, también, un propósito concreto: investigar las relaciones causa-efecto entre las condiciones manipuladas y los resultados medidos.

Las prácticas de investigación experimental se caracterizan, en primer lugar, por muchos tipos distintos de diseños experimentales de investigación. Aquí describiremos los tres tipos más comunes. El capítulo 8 nos presentará éstos y otros diseños con más detalle.

Experimental pura. La característica distintiva de una modalidad **experimental pura** es un diseño en el que se da una asignación al azar de los sujetos a los diferentes grupos. Con la **asignación aleatoria**, todos los sujetos tienen la misma oportunidad de estar en cada grupo. Este procedimiento, cuando se lleva a cabo con una muestra lo suficientemente grande, ayuda a asegurar que no existen grandes diferencias entre sujetos en cada grupo antes de que empiece el tratamiento experimental. Esto permite al investigador concluir que los resultados no se deben a diferencias entre los sujetos anteriores a recibir el tratamiento o mientras transcurre éste.

Las ciencias físicas y la biología utilizan frecuentemente diseños experimentales porque proporcionan la mayor aproximación para determinar el efecto de un factor sobre otro. En estas disciplinas, también es relativamente fácil encontrar las condiciones de la asignación aleatoria y la manipulación. Si un grupo de granjeros quiere determinar cuál de dos fertilizantes provoca mayor crecimiento,

pueden dividir grandes parcelas de tierra en secciones más pequeñas y fertilizar de forma aleatoria algunas secciones con el fertilizante A y las demás con el fertilizante B. Siempre y cuando la misma cantidad de lluvia y de sol y los mismos problemas de plagas y otros factores (lo que probablemente sería el caso) afecte a cada parcela, los granjeros pueden determinar qué fertilizante es mejor. En las ciencias sociales, sin embargo, especialmente en la educación, a menudo es difícil encontrar estas condiciones. Los experimentos son difíciles de adoptar, especialmente en investigación aplicada, en la que los investigadores minimizan los cambios a las condiciones que ocurren de forma natural.

Semiexperimental. La modalidad **semiexperimental** se encuentra próxima a la modalidad experimental. El propósito de los métodos es el mismo (determinar causa y efecto) y se da una manipulación directa de las condiciones. Sin embargo, no hay asignación aleatoria de sujetos. Una situación común de implementación de investigación semiexperimental implica varias clases o colegios en los que puede aplicarse para determinar el efecto de materiales curriculares o de métodos de enseñanza. Las clases están «intactas», organizadas ya para un propósito instructivo. Las clases no están asignadas de forma aleatoria y tienen profesores diferentes. Sin embargo, es posible administrar un tratamiento experimental a algunas de las clases y considerar a las otras como controles.

Caso único. La investigación en educación ha estado fuertemente influida por una tradición según la cual se estudian grupos de sujetos, en lugar de individuos. La razón para el estudio de grupos es que las diferencias entre los individuos y el error de medida pueden ser evaluados empleando la puntuación media del grupo completo. Sin embargo, en muchas situaciones, es imposible o inconveniente estudiar grupos enteros de sujetos. Además, el investigador puede estar interesado en uno o dos sujetos, no en grupos grandes. La **modalidad de investigación de caso único** ofrece una alternativa por la especificación de métodos que pueden ser empleados con un único individuo o con unos pocos y, aún así, permite unas conclusiones razonables de causa y efecto.

Comparable a la investigación semiexperimental, existe manipulación directa pero no hay asignación aleatoria. Suponga, por ejemplo, que está interesado en la efectividad de un nuevo programa de modificación del comportamiento para controlar la conducta desobediente. Hay tres alumnos en su clase con los que quiere probar el programa, por lo que un diseño de grupo resulta inapropiado. Concibiendo el problema como un diseño de caso único, primero debería registrar la conducta de uno de los alumnos para asegurarse de que la conducta es estable y consistente antes de la implementación del nuevo programa. Luego, continúa registrando la conducta después de que el programa haya sido implementado. Si observa un cambio en la conducta que coincide con la implementación del método nuevo y no pueden identificarse otras causas, podrá inferir que el programa ha provocado el cambio en la conducta.

Modalidades de investigación no experimentales

Las modalidades no experimentales describen alguna circunstancia que ha ocurrido o examinan las relaciones entre aspectos sin ninguna manipulación directa de las condiciones que son experimentadas. Existen cinco tipos de modalidades de investigación no experimentales: descriptivo, comparativo, correlacional, encuesta y *ex post facto*.

Descriptivo. La investigación que emplea una modalidad de investigación **descriptiva** refiere simplemente un fenómeno existente utilizando números para caracterizar individuos o un grupo. Evalúa la naturaleza de las condiciones existentes. El propósito de la mayoría de las investigaciones descriptivas se limita a caracterizar algo como es. Las siguientes cuestiones podrían ser contestadas a través de diseños descriptivos: ¿cuántas veces a lo largo del día usa la señora Jones el refuerzo negativo con sus alumnos? ¿Cuáles son las actitudes de los alumnos hacia la disciplina del colegio? ¿Cuál es el concepto de sí mismos de los alumnos de cuarto curso? ¿Cuál es el nivel de rendimiento lector de los alumnos en la escuela?

Comparativo. En una modalidad de investigación **comparativa**, el investigador estudia si existen diferencias entre dos o más grupos acerca del fenómeno que está siendo estudiado. Al igual que en la modalidad descriptiva, no se da manipulación o control directo de las condiciones, pero la aproximación comparativa se sitúa un paso más allá de la investigación descriptiva. Por ejemplo, más que describir simplemente las actitudes del alumno hacia la disciplina, un estudio comparativo podría investigar si las actitudes difieren según el curso o el género. Otro ejemplo podría ser comparar las notas de los deportistas y los que no hacen deporte. Las modalidades comparativas se emplean, a menudo, para estudiar las relaciones entre diferentes fenómenos, por ejemplo, la relación entre la participación en pruebas de atletismo y la nota media.

Correlacional. La investigación **correlacional** se ocupa de la valoración de las relaciones entre dos o más fenómenos. Este tipo de estudio, normalmente, implica una medida estadística del grado de la relación, llamado **correlación**. La relación medida es una afirmación sobre el grado de asociación entre las variables de interés. Una *correlación positiva* significa que los valores altos de una variable están asociados con los valores altos de una segunda variable. La relación entre altura y anchura, entre la puntuación de CI y de pruebas de rendimiento en un test y entre la autoestima y la notas son ejemplos de correlación positiva. Una relación o *correlación negativa* significa que los valores altos de una variable están asociados con los valores bajos de una segunda variable. Ejemplos de correlaciones negativas incluyen aquéllas entre ejercicio e infarto, entre realización exitosa de un test y sentimientos de incompetencia y entre absentismo escolar y rendimiento académico.

Encuesta. En la investigación mediante **encuesta** el investigador selecciona una muestra de sujetos y les administra un cuestionario o realiza entrevistas para recoger los datos. Las encuestas son utilizadas, frecuentemente, en la investigación educativa para describir actitudes, creencias, opiniones y otros tipos de información. Normalmente, la investigación está diseñada de modo que la información sobre un gran número de personas (la población) puede ser inferida de las respuestas obtenidas por un grupo más pequeño de sujetos (la muestra). Las encuestas se usan para una extensa variedad de propósitos. Pueden describir la frecuencia de características demográficas, explorar relaciones entre diferentes factores o esbozar las razones para prácticas particulares. En la investigación mediante encuesta se sigue una secuencia específica de pasos, que detallamos en el capítulo 7.

Ex post facto. La modalidad *ex post facto* se maneja para explorar las relaciones causales posibles entre variables que no pueden ser manipuladas por el investigador. El investigador diseña el estudio para cotejar dos o más muestras que son comparables, excepto en un factor determinado. Las causas posibles son estudiadas después de que hayan ocurrido. En lugar de manipular *lo que ocurrirá* a los sujetos, como en los diseños experimentales, la investigación se centra en lo que ha pasado diferencialmente a grupos de sujetos comparables; entonces examina si los sujetos de cada grupo son diferentes de alguna manera. Por ejemplo, una cuestión importante acerca del cuidado de los niños pequeños es el efecto relativo que puede tener el tipo de cuidado en la preparación para el colegio. Algunos programas son más «académicos» que otros. Dado que sería muy difícil manipular experimentalmente el tipo de programa al que un niño asiste, la modalidad de investigación *ex post facto* sería apropiada. El investigador podría identificar dos grupos de niños que poseen antecedentes similares pero que han estado expuestos a tipos de cuidado diferentes. A los sujetos se les debería administrar un test de preparación para el colegio para comprobar si los que han recibido un programa de cuidado altamente académico se diferencian de los niños que han asistido a un programa con orientación menos académica.

MODALIDADES DE INVESTIGACIÓN CUALITATIVA

Las modalidades de investigación cualitativas pueden ser clasificadas como interactivas o no interactivas. Estas modalidades cualitativas son importantes porque cada una tiene una historia prominente en alguna disciplina y ha generado revistas, libros y metodologías distintivas que caracterizan su aproximación. Como grupo, las modalidades cualitativas aparecen de forma frecuente en las publicaciones; exponen metodologías rigurosas y sistemáticas e ilustran la diversidad en el diseño de la investigación, el papel del investigador y las técnicas de recogida

de datos. Existen otras aproximaciones y variantes dentro de estas modalidades seleccionadas, pero su discusión está más allá del objetivo de este libro¹.

Investigación interactiva

La **investigación cualitativa** interactiva consiste en un estudio en profundidad mediante el empleo de técnicas cara a cara para recoger los datos de la gente en sus escenarios naturales. El investigador interpreta los fenómenos en términos de los significados que la gente les da. Los investigadores cualitativos construyen un dibujo holístico y complejo con descripciones detalladas de las perspectivas de los informantes. Algunos investigadores cualitativos tratan abiertamente los valores que forman la narrativa. Los investigadores interactivos describen el contexto del estudio, ilustran las diferentes perspectivas de los fenómenos y revisan continuamente cuestiones desde sus experiencias en el campo.

Las cinco modalidades de investigación interactiva de la tabla 2.1 son etnográfica (antropología y sociología), fenomenológica (psicología y filosofía), estudio de caso (ciencia social y humana y áreas aplicadas como evaluación), teoría fundamentada (sociología) y estudios críticos (varias disciplinas). Estas modalidades de investigación pueden ser organizadas desde (a) un enfoque sobre la experiencia individual vivida, como lo visto en fenomenología, estudio de casos, teoría fundamentada y algunos estudios críticos y (b) un enfoque sobre la sociedad y la cultura como el definido por la etnografía y algunos estudios críticos. Además, cada modalidad de investigación está apoyada en concepciones ya sean tradicionales o las más contemporáneas críticas o posmodernas. La investigación cualitativa «tradicional» asume que (a) una sociedad es una estructura ordenada y (b) el investigador aprende de los participantes manteniendo una posición relativamente neutral.

Etnografía. La **etnografía** es una descripción e interpretación de un grupo o sistema cultural o social. Sin embargo, aunque existen discrepancias considerables sobre el significado del término «cultura», el foco de interés se encuentra en los patrones aprendidos de acción, lenguaje, creencias, ritos y formas de vida. Como proceso, la etnografía implica un trabajo de campo prolongado, empleando, normalmente, la observación y entrevistas ocasionales con participantes de un grupo de actividad compartida y recogiendo productos grupales. Se emplea un estilo documental, enfocado en los detalles mundanos de la vida diaria y revelando las habilidades de observación del investigador. El punto de vista del informante es cuidadosamente expuesto a través de citas extensas y fieles para comunicar que lo que se describe no es el punto de vista del trabajador de campo, sino los

¹ Por ejemplo, ver N. Denzin y Y. Lincoln, eds. (2000), *Handbook of Qualitative Research and others*. Tratamos el constructivismo y la búsqueda naturalística como fundamental para la mayoría de investigaciones cualitativas interactivas.

comentarios auténticos y representativos de los participantes. El producto final es una descripción narrativa comprensiva y una interpretación que integra todos los aspectos de la vida del grupo e ilustra su complejidad.

Hay distintas variantes de la etnografía. Considerando que muchos antropólogos emplean la observación participativa en los estudios etnográficos de una cultura, los investigadores educativos utilizan la técnica para producir microetnografías (Erickson, 1973; Le Compte y Preissle, 1993; Wolcott, 1995). Una microetnografía es un estudio de observación participativa sobre un aspecto de un componente cultural (educación), como los participantes en una actividad educativa, por ejemplo, un aula urbana o los directores en un programa de innovación educativa. Si el estudio es un examen exhaustivo de una actividad muy específica o de una unidad muy pequeña dentro de una organización o del razonamiento práctico diario y la conversación, se denomina etnometodología (Schwandt, 1997).

Fenomenología. Se refiere tanto a una filosofía de la ciencia como a una modalidad de investigación. Un **estudio fenomenológico** describe los significados de una experiencia vivida. El investigador «suspende» o deja a un lado todos los prejuicios y recoge los datos sobre cómo los individuos descifran el significado de una experiencia o situación determinada. La meta de la fenomenología es transformar una experiencia vivida en una descripción de «su esencia, de tal forma que el efecto del texto es de inmediato un revivir reflexivo y la apropiación meditada de algo significativo» (Van Manen, 1990, p. 36). La técnica más característica es la entrevista extensa entre los informantes y el investigador, dirigida hacia la comprensión de las perspectivas de los informantes en su experiencia vivida diariamente con los fenómenos (Moustakas, 1994; Seidman, 1998). El estudio permite a los lectores sentir que pueden entender más ampliamente el concepto relacionado con una experiencia particular, tal como padecer una discapacidad física para un joven o la vivencia infantil para un niño de la separación y el divorcio de sus padres.

Estudio de caso. Mientras que algunos clasifican «el caso» como un objeto de estudio (Stake, 1995) y otros lo consideran una metodología (Yin, 1994), un **estudio de caso** examina un «sistema definido» o un caso en detalle a lo largo del tiempo, empleando múltiples fuentes de datos encontradas en el entorno. El caso puede ser un programa, un acontecimiento, una actividad o un conjunto de individuos definidos en tiempo y lugar. El investigador define el caso y su límite. Los casos no son elegidos por su representatividad: un caso puede ser seleccionado por su singularidad o puede ser utilizado para ilustrar un tema (Stake, 1995). El foco de interés puede ser una entidad (estudio intrasituación) o varias entidades (estudio multisituación). Sin embargo, cuantos más casos de situaciones individuales se añadan, tanto menor será la profundidad del análisis de una sola situación. El estudio proporciona una descripción detallada del caso, un análisis de los temas o asuntos y las interpretaciones o afirmaciones del investigador

sobre el caso. Estas interpretaciones pueden denominarse «lecciones aprendidas» (Guba y Lincoln, 1989).

Teoría fundamentada. Aunque el sello distintivo de la investigación cualitativa es la descripción detallada y el análisis de los fenómenos, la **teoría fundamentada** va más allá de la descripción para desarrollar conceptos «densos» (detallados) o declaraciones proposicionales condicionales que se relacionan con un fenómeno particular (Glaser y Strauss, 1967; Strauss y Corbin, 1998). El término «teoría fundamentada» a veces se usa de una forma no muy específica para referirse a cualquier aproximación para la formación de ideas teóricas que, de alguna manera, empieza con datos. Pero la metodología de la teoría fundamentada es un conjunto riguroso de procedimientos para elaborar una teoría sustantiva. Usando un método comparativo constante, el análisis de datos emplea simultáneamente técnicas de inducción, deducción y verificación. El investigador recoge, en primer lugar, los datos de la entrevista realizando múltiples visitas al campo. La recogida inicial de datos se efectúa para obtener una variedad de perspectivas sobre los fenómenos; luego, el investigador emplea la comparación constante para analizar a través de las categorías de información. Los datos se recogen hasta que las categorías de información están «saturadas». En este punto, el investigador selecciona los fenómenos centrales, desarrolla un «guión» y sugiere una matriz condicional que especifique las condiciones sociales e históricas y las consecuencias que influyen en el fenómeno.

Algunos investigadores consideran que la teoría fundamentada es una forma de inducción analítica modificada en la que la teoría inicial es desarrollada previamente como «hipótesis de trabajo» y luego «puesta a prueba» en los diferentes casos para desarrollar las propiedades que constituirán las proposiciones condicionales. En cada nueva visita, la recogida de datos está más limitada a la dimensión particular de la categoría de información. La inducción analítica modificada se utiliza a menudo en los estudios multisituación empleando la información del participante (Bodgan y Biklen, 1998).

Estudios críticos. Esta modalidad de investigación conforma la teoría crítica, la teoría feminista, la teoría de la raza y las perspectivas postmodernas, que asumen que el conocimiento es subjetivo². Estos investigadores también consideran la sociedad como esencialmente estructurada por clases y categorías, así como por raza, etnia, género y orientación sexual. De esta manera, una sociedad patriarcal mantiene la opresión de grupos marginados (Alter, 1991). Los investigadores críticos desconfían de la mayoría de las modalidades de investigación por ignorar las relaciones de poder implícitas en las técnicas de recogida de datos y por excluir otras formas de conocimiento (Marshall y Rossman, 1999). Mientras que la investigación

² Los estudios críticos, como grupo, emplean una variedad de metodologías incluyendo modalidades cuantitativas y cualitativas interactivas y no interactivas (Harding, 1987).

feminista y étnica enfoca el «problema» del estudio en el género y en la raza, las teorías postmodernista y crítica tienden a enfocarlo más sobre la sociedad y las instituciones sociales. Gran parte de la crítica de la investigación cualitativa tradicional está localizada en el análisis narrativo, la investigación-acción, la etnografía crítica, la investigación-acción participativa y la investigación feminista.

Los estudios críticos. Comienzan con un compromiso de exponer la manipulación social y la opresión, y de cambiar las estructuras sociales opresivas. A menudo tienen objetivos emancipatorios explícitos a través de la crítica sostenida o de la defensa directa y de la acción adoptada por el investigador o los participantes. Un investigador puede identificar su género, su edad, su raza, su situación social y sus posiciones políticas para informar al lector de que sus interpretaciones no están libres de valores (Carspecken, 1996; Rossman y Rallis, 1998).

Es necesario hacer dos aclaraciones. Primero, estas modalidades de investigación no son modalidades prudentes, aunque cada una posee implicaciones metodológicas diferenciadas. Las modalidades difieren en sus objetivos, la importancia de la teoría, la disciplina original, la técnica de recogida de datos, el papel del investigador y el formato narrativo del informe. Puesto que existe algún solapamiento entre estos cuatro usos cualitativos «tradicionales», se debe indicar la modalidad de investigación predominante para el estudio. Por ejemplo, tanto la etnografía de un sistema cultural como el estudio de caso de un sistema típico limitado investigan sistemas. Sin embargo, la etnografía (excepto la microetnografía) examina un sistema cultural; un estudio de caso está enfocado en unidades más pequeñas como un suceso o actividad y explora un rango de temas que no está limitado al comportamiento cultural, ritos o creencias (Creswell, 1998). De esta manera, la aproximación al estudio de los alumnos de un aula (microetnografía) puede ser un estudio de caso o etnográfico que explora un componente cultural (la educación). Un examen totalmente diferente de una clase particular podría ser un estudio crítico.

Segundo, usamos el término *diseño de estudio de caso* para referirnos a un «caso» estudiado en profundidad, en contraste con los diseños experimentales y no experimentales en los que los grupos son comparados a través de análisis estadísticos. En el diseño cualitativo, un «caso» es la entidad o fenómeno singular estudiado en profundidad. Podría tratarse de un individuo, un grupo de alumnos, un colegio, un programa o un concepto. El diseño de estudio de caso se aplica a todos los tipos de investigación, tradicional y no tradicional, e interactiva y no interactiva. Por lo tanto, metodología de estudio de caso no es lo mismo que diseño de estudio de caso.

Investigación no interactiva

Las modalidades de investigación no interactivas, conocidas como **investigación analítica**, estudian conceptos y sucesos históricos a través de un análisis de

documentos. El investigador identifica, estudia y, luego, sintetiza los datos para proporcionar un conocimiento del concepto o del suceso pasado que puede o no haber sido directamente observable. Los documentos autenticados son la fuente principal de datos. El investigador interpreta «hechos» para proporcionar explicaciones del pasado y descubre los significados educativos colectivos que pueden subyacer a prácticas y temas actuales.

Ejemplos de investigación analítica incluyen el análisis de conceptos y el análisis histórico. El **análisis de conceptos** es el estudio de conceptos educativos como «aprendizaje cooperativo», «agrupamiento por habilidades», «liderazgo» para describir los diferentes significados y el uso apropiado del concepto. El análisis histórico implica una recogida sistemática y una crítica de documentos que describen sucesos pasados. Los historiadores educativos estudian programas educativos, prácticas, instituciones, personas, políticas y movimientos pasados. Éstos normalmente se interpretan en el contexto de tendencias históricas, económicas, sociales, militares, tecnológicas y políticas. Los análisis examinan causas y las consecuencias subsiguientes y, a menudo, describen el pasado de los acontecimientos actuales.

TÉCNICAS DE RECOGIDA DE DATOS

Otra forma de clasificar los tipos de investigación es examinar la técnica empleada en el estudio para la recogida de datos. Es decir, ¿cómo obtuvo los datos el investigador? Al igual que las modalidades de investigación, las técnicas pueden clasificarse como cuantitativas o cualitativas como se indica en la tabla 2.2. Una diferencia fundamental es que las aproximaciones cuantitativas utilizan números para describir el fenómeno, mientras que las técnicas cualitativas recurren a descripciones narrativas (sin embargo, algunos estudios emplean técnicas de recogida de datos cualitativas y, sin embargo, emplean números para resumir los resultados). Aunque la mayoría de las técnicas se usan con cualquiera de las modalidades de investigación, los diseños de investigación más específicos están fuertemente asociados a una técnica. Virtualmente, todos los diseños experimentales emplean técnicas cuantitativas y la mayoría de los diseños de estudios de caso utilizan técnicas cualitativas. Más adelante encontrará una introducción a cada técnica. Se estudiarán con más detalle en los siguientes capítulos.

Técnicas cuantitativas

Las técnicas de investigación cuantitativa destacan categorías *a priori* para la recogida de datos en forma de números. La meta es proporcionar descripciones estadísticas, relaciones y explicaciones. Las técnicas cuantitativas se emplean en diseños experimentales, descriptivos y correlacionales, como una forma de

TABLA 2.2: TÉCNICAS DE RECOGIDA DE DATOS

	Cuantitativa	Cualitativa
Tipos	Observaciones estructuradas. Entrevistas estandarizadas. Tests de lápiz y papel. Cuestionarios. Evaluaciones alternativas.	Observación participante. Observaciones de campo. Entrevistas en profundidad. Documentos y artefactos. Técnicas suplementarias.
Características	Utilizan instrumentos para la recogida de datos. Los datos son números. Decisión <i>a priori</i> en la presentación de los datos. Los datos toman una forma –la respuesta está determinada por la prueba. Los datos se tabulan y describen estadísticamente. El significado se deriva de los procedimientos estadísticos empleados.	Los datos se recogen sin instrumentos. Los datos son palabras. No hay decisión <i>a priori</i> con respecto a la presentación de los datos; depende de los datos recogidos. Los datos toman muchas formas –notas de campo, documentos, notas de entrevista, grabaciones, artefactos. La tabulación se limita a ayudar a identificar patrones; usados como apoyo de los significados cualitativos. El significado se deriva de las estrategias cualitativas empleadas.

resumir un número elevado de observaciones y de indicar numéricamente la cantidad de error en la recogida y presentación de los datos.

Observaciones estructuradas. Cualquier investigación requiere algún tipo de observación sobre las personas, cuestiones o procesos. Sin embargo, lo que nos interesa de las **observaciones estructuradas** es un tipo particular de recogida de datos, en el que el investigador observa directamente, auditiva y visualmente, algún fenómeno y, luego, registra de forma sistemática las observaciones resultantes. El observador ha predeterminado categorías específicas de conducta que serán registradas; lo que observará se determina antes de que la investigación se realice. Normalmente, se identifican las unidades de conducta y se emplea un procedimiento sistemático para comprobar o cuantificar conductas específicas. Por ejemplo, el observador puede registrar cuántas veces hacen preguntas los alumnos, qué tipo de preguntas realizan o cuánto tiempo emplea el profesor para responder cada pregunta.

Entrevistas estructuradas. En una entrevista existe interacción verbal directa entre el entrevistador y el sujeto. Una **entrevista estructurada** es un conjunto estándar de preguntas, que se plantean de forma oral y en persona y que están

preparadas de antemano. Normalmente, las preguntas dan a elegir una respuesta o son semiestructuradas. Cuando se plantea una pregunta en la que el sujeto puede elegir la respuesta, éste la seleccionará de las alternativas proporcionadas por el entrevistador. Las preguntas semiestructuradas están redactadas para permitir respuestas únicas para cada sujeto. Independientemente del tipo de pregunta, las respuestas se codifican, tabulan y resumen numéricamente.

Cuestionarios. Los cuestionarios abarcan una variedad de documentos en los que el sujeto responde a cuestiones escritas que sonsacan reacciones, opiniones y actitudes. El investigador elige o construye un conjunto de preguntas adecuadas y le pide al sujeto que las conteste, normalmente en forma de preguntas en las que debe elegir la respuesta (por ejemplo: sí, no, quizás). Esta técnica de recogida de datos es muy habitual en la investigación educativa y muchos sondeos mediante encuesta emplean cuestionarios. Los cuestionarios no son necesariamente más fáciles que otras técnicas y se deberían emplear cuidadosamente.

Tests de lápiz y papel. En un **test de lápiz y papel** al entrevistado se le plantean una serie de preguntas que están puntuadas objetivamente. Los formatos típicos de los ítems incluyen elección múltiple, emparejamiento, verdadero y falso y completado. Las puntuaciones resultantes de los tests se utilizan como datos. Debido a que estos tipos de tests están bien establecidos y poseen elevadas cualidades técnicas, se usan a menudo en investigación educativa como una medida del rendimiento de los alumnos. Sin embargo, como veremos en capítulos posteriores, se necesita precaución cuando se emplean estos tipos de medidas.

Evaluaciones alternativas. Las **evaluaciones alternativas** son medidas de rendimiento que requieren la demostración de un nivel de eficiencia al tener el entrevistado que crear, producir o hacer algo. Un tipo de evaluación alternativa está basado en el rendimiento, tal como hacer una presentación oral, escribir un proyecto, hacer una presentación musical, demostraciones, rendimiento atlético y otros proyectos. Las carpetas de trabajo constituyen otro tipo de evaluación alternativa. Muchas evaluaciones alternativas son *auténticas*, reflejan los problemas y las situaciones de la vida real. Aunque las evaluaciones alternativas se han hecho populares como una técnica empleada en investigación en los últimos años, están llenas de dificultades técnicas. Esto es debido, principalmente, a la naturaleza subjetiva de la puntuación del rendimiento o del resultado.

Técnicas cualitativas

Como hemos indicado anteriormente, las técnicas cualitativas recogen los datos principalmente en forma de palabras en lugar de números. El estudio proporciona una descripción narrativa detallada, un análisis y una interpretación de los fenómenos. La mayoría de los investigadores cualitativos emplean varias técnicas

en sus estudios pero, normalmente, seleccionan una como método central. Hasta cierto punto, la observación participante, la observación y la entrevista forman parte de cualquier investigación interactiva. Se recurre a otras estrategias para complementar o incrementar la credibilidad de los hallazgos. La investigación no interactiva depende principalmente de documentos. Las técnicas cualitativas proporcionan descripciones verbales para retratar la riqueza y complejidad de los acontecimientos que ocurren en escenarios naturales desde la perspectiva de los participantes. Una vez recogidos, los datos son analizados inductivamente para generar descubrimientos.

Observación participante. La observación participante es una técnica interactiva de «participar» hasta cierto punto en las situaciones que ocurren, de forma natural, durante un periodo de tiempo y escribir extensas notas de campo que describen lo que ocurre. El investigador no recoge datos para responder a una hipótesis específica; más bien, las explicaciones se derivan inductivamente de las notas de campo. Ya que el contexto de las observaciones es importante, el investigador debe ser cuidadoso en documentar su papel en la situación y qué efecto puede tener en los resultados. La mayoría de los trabajadores de campo permanecen a una distancia respetuosa de los informadores, cultivando empatía pero no simpatía, compenetración pero no amistad, familiaridad pero nunca «convirtiéndose en nativo» (Schwandt, 1997). Sin embargo, otros investigadores, debido a que pasan un tiempo considerable en ese entorno, también aprenden de su propia experiencia en ese entorno. Estas reflexiones personales se integran en los análisis procedentes del grupo social de interés. La investigación colaborativa y participante introduce la noción de participación activa por parte del investigador y el compartimiento del papel de investigador con los participantes. En todas las variantes de la observación participante, el papel de la investigación se establece al principio del estudio y, luego, se revisa con documentación.

Observación de campo. Una técnica fundamental para la mayoría de la investigación cualitativa es la observación de campo (informes directos y presenciales de la acción social cotidiana y de los escenarios en forma de notas de campo). Las observaciones de campo cualitativas son descripciones detalladas de sucesos, personas, acciones y objetos en escenarios. La observación de campo se utiliza en la recogida de datos interactiva, como en la observación participante y en las entrevistas en profundidad. En la primera, el investigador confía en la observación cuidadosa cuando explora inicialmente varias áreas de interés en un lugar, seleccionando las que hay que estudiar en detalle y buscando patrones de conducta y relaciones. En la segunda, el investigador anota el lenguaje corporal no verbal y las expresiones faciales del entrevistado como ayuda para interpretar los datos verbales.

Entrevistas en profundidad. La **entrevista en profundidad**, a menudo, se caracteriza por una conversación con un objetivo. El investigador puede emplear

una guía de entrevista general o un protocolo, pero no un conjunto de preguntas específicas que el mismo ha formulado para cada entrevista. Mejor dicho, hay unas pocas preguntas generales, con una libertad considerable para seguir una gama amplia de temas. El entrevistado puede dar forma al contenido de la entrevista centrándose en temas de importancia o interés. De hecho, el investigador normalmente anima a la persona a hablar con detalle sobre las temas de interés. Una entrevista en profundidad, por lo general, dura como mínimo una hora. El investigador, a menudo, graba las entrevistas y, posteriormente, las transcribe para analizar los temas frecuentes de las descripciones de experiencias.

Documentos y enseres. Los **documentos** son registros de sucesos pasados que han sido escritos o impresos; pueden ser notas anecdóticas, cartas, diarios y documentos. Los documentos oficiales incluyen papeles internos, comunicaciones a varias personas, archivos de estudiantes y de personal, descripciones de programas y datos estadísticos institucionales. En las técnicas de recogida de datos interactivas, el investigador encuentra esta documentación en el lugar o bien un participante le ofrece compartir dicha información. Los documentos también proporcionan información de los antecedentes del tema.

Los documentos son la fuente de datos más importante en el análisis conceptual y en los estudios históricos. Normalmente, se catalogan y conservan en archivos, almacenes de manuscritos o bibliotecas. Estas colecciones están dispersas a través de los Estados Unidos. Debido a que las colecciones recogen documentos originales, no se permite que abandonen el archivo. Primero, el historiador debe localizar los documentos, para lo que usa índices de archivos y, después, debe obtener un permiso para poder trabajar con los textos originales. Para asegurar la autenticidad de los documentos se les aplican técnicas rigurosas de crítica. Otros documentos están en colecciones privadas conservadas por los individuos o miembros de su familia; para localizar estos documentos es necesario aplicar los mismos procedimientos metodológicos.

Los **enseres** son objetos materiales y símbolos de un suceso, grupo, persona u organización pasados o presentes. Estos objetos son entidades tangibles que revelan procesos, significados y valores sociales. Los logotipos y las mascotas de los colegios son ejemplos de símbolos; los diplomas, premios y los productos de los estudiantes como dibujos, trabajos y murales son ejemplos de objetos. El investigador cualitativo está más interesado en los significados asignados a los enseres y en los procesos sociales que los producen que en los objetos en sí mismos.

Técnicas suplementarias. Los investigadores cualitativos emplean una gran variedad de técnicas suplementarias para obtener hallazgos creíbles. Las técnicas suplementarias son aproximaciones seleccionadas para ayudar a interpretar, elaborar o corroborar los datos obtenidos desde la observación participante, la entrevista en profundidad, los documentos y los enseres. Aunque cada una de estas técnicas es un método completamente independiente, se incorporan a menudo

dentro de una investigación cualitativa. Algunos ejemplos de técnicas suplementarias son: técnicas visuales, películas etnográficas y el uso de vídeos, películas y fotografías para captar la vida cotidiana de un grupo de estudio. La quinesia es el estudio de los movimientos corporales y de la comunicación no verbal. El análisis del espacio simbólico interpersonal se denomina proxémia. Un grupo focal es una entrevista grupal a ocho de quince individuos seleccionados que comparten ciertas características relevantes para el propósito del estudio. Las encuestas pueden ser un instrumento construido por los participantes o la confirmación para el investigador de los resultados.

CÓMO LEER INVESTIGACIÓN

La investigación puede presentarse de muchas formas, la más habitual suele ser un artículo publicado o una ponencia enviada a un congreso. El propósito del informe es indicar claramente el trabajo del investigador, por qué lo ha realizado y cuál es su significado. Para hacerlo con efectividad, los investigadores utilizan un formato más o menos estándar. Este formato es similar al proceso de conceptualización y realización de la investigación. Dado que el proceso de investigación es distinto en la aproximación cuantitativa y cualitativa, existen diferencias en los formatos de presentación que se emplean para cada enfoque. Vamos a repasar, de forma separada, los formatos básicos para cada tipo de aproximación.

En la lectura de la investigación es importante juzgar la credibilidad global del estudio. Este juicio está basado en una evaluación de cada una de las secciones principales del informe. Cada parte del informe contribuye a la credibilidad global del estudio. De esta manera, siguiendo una descripción del formato de cada tipo de investigación presentaremos directrices que serán útiles para evaluar cada sección del informe.

Cómo leer una investigación cuantitativa: un ejemplo no experimental

Aunque no existe un formato de presentación de investigación cuantitativa universalmente aceptado, la mayoría de los estudios se rigen por la secuencia de búsqueda científica. Se dan algunas variaciones en los términos empleados, pero los componentes citados a continuación aparecen en la mayoría de los estudios.

1. Resumen
2. Introducción
3. Enunciado del tema de investigación
4. Revisión bibliográfica
5. Enunciado de las hipótesis o preguntas de investigación

6. Metodología
 - a. sujetos
 - b. instrumentos/ pruebas
 - c. procedimiento
7. Resultados
8. Exposición, implicaciones, conclusiones
9. Referencias

Al escribir un informe de investigación, el autor comienza con la introducción y continúa de forma secuencial hasta la conclusión. Al planificar la realización de una investigación, los investigadores comienzan por formular un tema de investigación. El extracto 2.1 ilustra sobre los componentes citados anteriormente en un artículo de investigación cuantitativa.

Resumen. El *resumen* (a veces se emplea el término sin traducir *abstract*) es un párrafo que resume el artículo. Sigue al nombre de los autores y, normalmente, tiene un tipo de letra en cursiva o más pequeña que el resto del artículo. La mayoría de los resúmenes contienen un enunciado del propósito del estudio, una breve descripción de los sujetos y lo que hicieron durante el estudio y un resumen de los resultados más importantes. El resumen es útil porque proporciona una visión general de la investigación y, tras leerlo, el lector sabrá si leer el artículo completo.

Introducción. La *introducción*, normalmente, se limita al primero o a los dos primeros párrafos del artículo. El propósito de la introducción es contextualizar el estudio. Esto se logra citando la investigación anterior sobre el tema general, citando investigadores destacados en el área o desarrollando el contexto histórico del estudio. La introducción actúa como una presentación inicial del propósito más específico del estudio. En el extracto 2.1 la introducción se localiza en los dos primeros párrafos.

Tema de investigación. El primer paso, al planificar un estudio cuantitativo, es formular un tema de investigación. El *tema de investigación* consiste en un enunciado claro y breve que expresa el propósito del estudio. Los investigadores comienzan con una idea general de lo que pretenden estudiar, como la relación de la autoestima y el rendimiento y, entonces, rematan su meta general con una concisa frase que expresa de forma más concreta el asunto que se está investigando. Por ejemplo, ¿qué relación existe entre la autoestima de los alumnos de cuarto curso sobre su capacidad en matemáticas y su rendimiento en las clases de matemáticas que indican las puntuaciones de los tests estandarizados?

El enunciado del tema de investigación se puede encontrar en varias localizaciones dentro del artículo. Puede ser la última frase de la introducción o puede seguir a la revisión bibliográfica e ir justo antes de la sección de metodología. En nuestro artículo ilustrativo, el tema aparece en el primer párrafo.

EXTRACTO 2.1**EJEMPLO DE UN INFORME DE INVESTIGACIÓN CUANTITATIVA****RELACIÓN ENTRE *LOCUS* DE CONTROL, NIVEL ACADÉMICO Y GÉNERO EN LOS ALUMNOS DE ENSEÑANZA SECUNDARIA**

*Marvin W. Boss, University of Ottawa
y Maurice C. Taylor, Algonquin College*

Se investigaron las relaciones entre el locus de control, el programa académico y el género de los alumnos de enseñanza secundaria de noveno curso. Se administraron la forma modificada de la escala de locus de control y el cuestionario Intellectual Achievement Responsibility Questionnaire a 267 alumnos de bachillerato pertenecientes a programas de diferentes niveles. Como se había supuesto, los alumnos del nivel avanzado tenían más control interno que los de nivel general o básico. Igualmente, los alumnos de nivel avanzado eran más responsables de sus fallos académicos que los alumnos de nivel general. También se discuten las diferencias de género que se relacionan con expectativas específicas de las situaciones de rendimiento académico.

El propósito de este estudio era determinar si el nivel académico y el género de los alumnos de enseñanza secundaria de noveno curso están relacionados con las expectativas generalizadas para un locus de control de refuerzo. Además se indagó sobre la relación entre el nivel académico y el sexo con las expectativas específicas en situaciones de rendimiento académico.

El *locus* de control se define (Rotter, 1966) como una expectativa generalizada del grado en el que una persona percibe que lo que le ocurre en la vida es consecuencia de su conducta. Las personas definidas como «internas» creen que ejercen más control sobre lo que les ocurre y les afecta. En contraste, los «externos» tienden a creer que tienen poco control sobre lo que les ocurre. Percibimos estas expectativas como resultado de muchas experiencias pasadas.

Lefcourt (1980) sugiere que algunas expectativas son muy generales y se relacionan con la mayoría de las situaciones de la vida; otras son más específicas y se relacionan con situaciones de la vida más concretas. La mayoría de los instrumentos que se manejan para medir el *locus* de control proporcionan medidas de expectativa generalizada. El cuestionario *Intellectual Achievement Responsibility (IARQ)* (Crandall, Katlovsky y Crandall, 1965) proporciona una medida de expectativas específicas de responsabilidad para el éxito y el fracaso académico. La escala da una puntuación para el éxito internalizado (I+) y una para el fracaso internalizado (I-).

En una revisión bibliográfica sobre la relación entre el *locus* de control (expectativas generalizadas y específicas) y el rendimiento, Bar-Tal y Bar-Zohan (1977) afirmaron que 31 de los 36 estudios revisados indicaban un relación significativa entre el *locus* de control y el rendimiento académico; además, los internos tenían un rendimiento más alto que los externos. McGhee y Crandall (1968) investigaron las expectativas específicas y establecieron I+ como predictor del rendimiento masculino I- como predictor del femenino.

Nowicki y Strickland (1973) encontraron que, de forma particular para los hombres, una puntuación interna en la escala de Nowicki-Strickland se relaciona con la competencia académica y la madurez social y parece haber una correlación con conductas de independencia, competición y automotivación.

Dweck y Licht (1980) mantienen que los chicos y chicas tienen maneras características diferentes de hacer frente a los resultados positivos y negativos. Ambos sexos interpretan sus éxitos y sus fracasos de manera diferente y tienen puntos de vista diferentes de las implicaciones de sus habilidades. Se diferencian en la persistencia en sus intentos para solucionar un problema difícil, en la calidad de su actuación después del fracaso y en sus elecciones de tarea cuando encuentran una dificultad.

Lochel (1983) revisó las diferencias de sexo en el rendimiento. Sugiere que las mujeres están más inclinadas a hacerse responsables de los fallos. Las ve con escasa confianza en sus habilidades y sin estar preparadas para hacer frente al fracaso.

Tal y como ha apuntado recientemente Parsons (1981) la conclusión de que los hombres tienden a atribuir sus fallos a causas externas o inestables mientras que las mujeres tienden a atribuir los fracasos a causas internas parece ser una simplificación excesiva. Cita varios ejemplos en una revisión de la bibliografía sobre atribuciones. Usando el IARQ, Dweck y Reppucci (1973) informaron que no hay diferencias de sexo en la atribución del fracaso a factores internos, pero encontraron que era ligeramente más probable que los chicos atribuyeran sus fracasos a la falta de esfuerzo que las chicas. En contraste, Crandall y otros (1965) encontraron chicas que interiorizaban más sus errores; Beck (1977) no obtuvo diferencias de sexo, ni internamente ni en la falta de atribuciones de esfuerzo; Diener y Dweck (1978) no hallaron diferencias significativas en falta de esfuerzo o atribución interna para los fracasos debidos al sexo; y Nicholls (1975) no encontró el efecto principal para la diferencia de sexo en las atribuciones de fracaso debido a la falta de esfuerzo. De forma similar, para las medidas de atribuciones de errores a causas externas emergen patrones inconsistentes (Dweck y Reppucci, 1973; Nicholls, 1975; Parsons, 1978).

Los alumnos de secundaria de Ontario dan clase en uno de estos tres niveles: avanzado, general y básico. Se espera que los alumnos de nivel avanzado continúen su educación en la universidad. Los alumnos del nivel general están preparados para continuar su educación con cursos técnicos o profesionales relacionados con profesiones específicas o pueden no continuar sus estudios. Los alumnos de nivel básico siguen su educación con cursos vocacionales en la escuela secundaria y, de forma general, no se les considera capaces de llegar con éxito a los niveles general y avanzado.

Los alumnos matriculados en los programas vocacionales de nivel básico son dirigidos a esos programas por sus profesores de educación primaria. Tradicionalmente, estos estudiantes exhiben un registro de rendimiento bajo y muchos tienen problemas específicos de aprendizaje. Además, es probable que hayan recibido ayuda especial durante su estancia en la escuela elemental.

Los estudiantes, con los consejos de sus profesores de primaria y secundaria, eligen estudiar cualquiera de los cursos académicos de nivel general o avanzado (biología, química, inglés, geografía, historia, matemáticas, física). Para tomar esta decisión, el

rendimiento en el pasado es probablemente más determinante que su aptitud. Sin embargo, tanto el rendimiento como la aptitud deberían ser, sin lugar a dudas, más bajos entre los alumnos de nivel general.

Debido a estas experiencias pasadas que influyen en que los estudiantes seleccionen o sean orientados a niveles específicos, se estableció la hipótesis de que los alumnos de nivel avanzado serían los más internos y los alumnos de nivel básico los menos internos para las medidas de *locus* de control, tanto generales como específicas.

Siguiendo la bibliografía revisada, también se estableció la hipótesis de que las mujeres asumirían más responsabilidad por sus fracasos que los hombres.

Metodología

Sujetos de investigación. Los sujetos de la investigación fueron 267 alumnos de noveno curso de tres escuelas de secundaria, localizadas en una comunidad suburbana de clase alta. Uno de los colegios era un centro vocacional, admitía alumnos de todo el distrito y sólo tenía alumnos de nivel básico. Los otros dos colegios (académicos) admitían alumnos de los niveles general y avanzado y se consideraba que tenían alumnos representativos del distrito. Si los alumnos habían cursado dos o más cursos académicos del nivel general eran clasificados como generales.

Instrumentos de medida. Se empleó la escala para niños del *locus* de control de Nowicki-Strickland (NS) para medir las expectativas generalizadas y el IARQ se empleó para medir las expectativas específicas (I+ y I-). Debido al tiempo reducido de comprobación se utilizó una versión de NS de 21 ítems para los cursos 7^o-12^o (Nowicki y Strickland, 1973) y una versión de 20 ítems del IARQ (Crandall, 1968) para los cursos 6^o-12^o. Crandall (1968) informó de correlaciones de .89 para I+ y de .88 para I- entre la forma reducida (cursos 6^o-12^o) y la forma extendida del IARQ.

Recogida de datos. Las pruebas de medida fueron administradas a los alumnos de noveno curso por sus profesores de inglés en tiempo de clase. El investigador se reunía con los profesores o con el subdirector antes de la entrega de los tests para repasar los procedimientos de administración. La prueba se completó en diciembre de 1985 en dos colegios (el vocacional y uno de los académicos). Los datos del otro colegio académico se recogieron en abril de 1986¹.

Resultados

Se usó el coeficiente α para estimar la fiabilidad de las tres medidas. Los resultados fueron .70 para el NS, .43 para I+ y .46 para I-. La fiabilidad para I+ y I- era bastante baja; lo que podría haber incrementado la probabilidad de errores de tipo II cuando se analizaron los datos.

Antes de comprobar las hipótesis, se empleó el análisis de la varianza para comprobar las diferencias entre las dos escuelas académicas. No se encontraron diferencias significativas entre ellas.

¹ Originalmente, los datos se recogieron del segundo colegio académico en diciembre de 1985. Sin embargo, se desecharon los datos porque no se siguieron procedimientos uniformes de administración. Por eso se incluyó otra escuela.

Para determinar si las diferencias eran las establecidas como hipótesis, se empleó el análisis de varianza con nivel y sexo como variables independientes y las puntuaciones NS, I+ y I- como variables dependientes. Los resultados se muestran en la tabla 1.

TABLA 1: Resultados de análisis de varianza con nivel y sexo como variables independientes y puntuaciones NS, I+ y I- como variables dependientes

Variable dependiente	Fuente	SC	gl	MC	F
NS	Nivel	212,45	2	106,23	8,11*
	Sexo	27,12	1	27,12	2,07
	Interacción	14,49	2	7,25	0,55
	Error	2999,67	229	13,10	
I+	Nivel	6,21	2	3,10	0,97
	Sexo	5,41	1	5,41	1,69
	Interacción	6,79	2	3,40	1,06
	Error	732,11	229	3,20	
I-	Nivel	27,74	2	13,88	3,78*
	Sexo	6,77	1	6,77	1,85
	Interacción	5,07	2	2,54	0,69
	Error	840,17	229	3,67	

*Significativo en nivel 0.05

Ya que los tamaños muestrales no eran iguales, se usó un análisis de medias sin ponderar.

Se encontraron diferencias significativas sólo para el nivel en el Nowicki-Strickland y en el I-. *Post hoc*, la modificación Tukey-Kramer del procedimiento WSD de Tukey se utilizó para determinar qué grupos diferían significativamente. Para el NS, los alumnos de nivel avanzado eran significativamente más internos que los alumnos de los niveles general o básico (ver tabla 2 para las medias). Estos resultados fueron los esperados. Para I-, los alumnos de nivel avanzado eran significativamente más internos que los alumnos de nivel general; los alumnos de nivel básico no diferían significativamente de los de nivel general o avanzado.

No se encontraron diferencias de sexo para ninguna de las variables dependientes. Se había pronosticado que las mujeres tendrían puntuaciones I- mayores que los hombres. Los resultados encontrados iban en la dirección contraria, si bien no resultaron significativos. Los hombres tuvieron una media I- mayor que la de las mujeres pero su media I+ fue menor que la de las mujeres. Para profundizar en estos resultados, se empleó la puntuación diferencial I+ menos I- como variable dependiente. Se encontró una diferencia significativa en el sexo. Para los hombres, la diferencia entre las puntuaciones I+ y I- fue menor ($x = .32$) que para las mujeres ($x = .88$). Para continuar la exploración, se usaron pruebas *t* para determinar si los hombres o las mujeres puntuaban de forma significativa más en I+ que en I-. Las mujeres puntuaron significativamente más alto en I+ que en I- ($t = 3.35$); no había una diferencia significativa para los hombres ($t = 1.68$).

Los resultados como consecuencia del sexo no fueron consistentes con lo que se había sugerido en la bibliografía. Examinando las medias para sexo por grupos, parece que las diferencias de sexo son atribuibles a alumnos de los niveles básico y general. En el nivel avanzado, las diferencias de medias entre I+ y I- fueron .46 para los hombres y .31 para las mujeres, mientras que en el nivel general las diferencias fueron .26 para los hombres y 1.43 para las mujeres; en el nivel básico los hombres se diferenciaban en .14 y las mujeres en 1.27. De esta manera, para los niveles básico y general, las mujeres obtuvieron puntuaciones en atribución interna sustancialmente más altas para el éxito académico que para el fracaso. Sería interesante explorar estas relaciones en estudios futuros.

TABLA 2: Medias y desviaciones típicas por nivel y sexo para las puntuaciones NS, I + y I-

	HOMBRE			MUJER			TOTAL	
	n	x	s	n	x	s	n	x
Nowicki-Strickland								
Básico	30	9,97	3,09	15	10,33	3,18	45	10,09
General	50	8,76	3,28	33	9,21	4,21	83	8,94
Avanzado	65	6,86	3,50	42	8,31	3,82	107	7,43
Total	145	8,16		90	8,98			
I+								
Básico	30	6,67	1,58	15	7,27	2,40	45	6,87
General	50	6,30	1,94	33	6,85	1,55	83	6,52
Avanzado	65	6,97	1,69	42	6,83	1,72	107	6,92
Total	145	6,68		90	6,91			
I-								
Básico	30	6,53	2,09	15	6,00	2,54	45	6,36
General	50	6,00	1,94	33	5,42	2,23	83	5,80
Avanzado	65	6,36	1,64	42	5,80	1,63	107	6,51
Total	145	6,35		90	6,03			

Exposición

Los resultados sugieren que la expectativa generalizada de refuerzo está relacionada con el nivel del programa académico de los alumnos de secundaria de noveno curso de Ontario. El alumno del programa de nivel avanzado tiene más control interno que el de los niveles general y básico. Se asume que los alumnos de nivel avanzado tienen un mayor rendimiento que los de nivel general y básico. Aunque la relación entre *locus* de control y logro académico ha sido constatada con anterioridad, el hecho de que se produzca en el presente estudio, acentúa la importancia de la expectativa generalizada para determinar el rendimiento académico. Los resultados de esta investigación apoyan la creencia de que cuanto más interna es la orientación del individuo, mayor es su aprovechamiento.

Se encontró, asimismo, una relación entre las expectativas específicas y el tipo de programa académico de los alumnos de secundaria. Los alumnos de nivel avanzado se atribuían más responsabilidad interna de sus fracasos académicos o intelectuales que los

alumnos de nivel general. Sorprendentemente, ni los alumnos de nivel general ni del nivel avanzado mostraban más responsabilidad interna respecto a sus fallos académicos que los alumnos de nivel básico. Bien puede suceder que los alumnos de las escuelas vocacionales funcionen mejor como un grupo de estudio separado antes que integrado. Los alumnos de un centro vocacional pueden tener una cultura y un conjunto de convicciones sobre responsabilidad para fracasos académicos ligeramente distinto al que podrían tener en un colegio académico. Aislados de los alumnos de nivel general y avanzado, el alumno de nivel básico puede sentirse más aceptado y es más probable que persista ante el fracaso temporal. Los alumnos de nivel general educados en el mismo entorno que los alumnos de nivel avanzado, sin lugar a dudas, reciben mensajes del entorno que sugieren que no son capaces de decidir sus propios refuerzos. Como se ha sugerido arriba, esto puede estar en función de un rendimiento inferior durante la escuela primaria. En una investigación más detallada con los tres grupos de alumnos, se deberían considerar las relaciones de responsabilidad de los éxitos y fracasos intelectuales-académicos con varios factores personales y diferentes métodos didácticos en el aula.

Los resultados del estudio señalan, también, el problema del entrenamiento en atribuciones de los alumnos con bajo rendimiento. Como apuntaron Dweck y Licht (1980), los cocimientos de los alumnos orientados a la universidad (una orientación motivacional) reflejan su tendencia a mirar hacia el futuro, resaltar lo positivo e invertir sus energías en seguir activamente las estrategias pertinentes de solución (p.201). Aceptado esto, ¿no parece una estrategia razonable enseñar a los alumnos con indefensión a atribuir sus fracasos a factores variables como el esfuerzo personal? De acuerdo con Dweck y Licht (1980) podría parecer necesario librar a los individuos con una indefensión seria de sus cualidades de inadaptación mediante un método directo antes de que puedan emplear sus propias autoinstrucciones de mayor adaptación y el autocontrol del estudiante orientado hacia la universidad.

Este estudio también apoya la idea de que, aunque el *locus* de control parezca relacionado al nivel académico, su relación con el sexo no es consistente. La hipótesis referente al sexo se pudo confirmar. Sin embargo, las mujeres tenían más capacidad para responsabilizarse de los éxitos y menos para aceptar su responsabilidad respecto a los fracasos. Las diferencias entre las puntuaciones I+ y I- mostraron que los hombres ($x = .33$) tenían una diferencia significativamente menor que las mujeres ($x = .88$). Además, las mujeres eran significativamente más internas para los éxitos que para los fracasos. Estas diferencias mostraron ser el resultado de una diferencia bastante grande entre las puntuaciones I+ y I- de las alumnas de nivel básico y general. Estos resultados contrastan con los de Lochel (1983), quien sugirió que las mujeres están más inclinadas a responsabilizarse de los fracasos que de los éxitos. ¿Podría ser que las alumnas de los niveles básico y general tengan una necesidad mayor de responsabilizarse de los éxitos y nieguen su responsabilidad ante los fracasos que las alumnas de nivel avanzado? Debido al comportamiento social, los niños con fracaso escolar reciben demasiado refuerzo negativo. Además, reciben este refuerzo por su comportamiento académico. Así, puede que no se sintieran tan amenazados por asumir la responsabilidad de sus fracasos, ya que puede que no diferenciaron entre los refuerzos negativos procedentes de los comportamientos sociales de los del rendimiento académico.

Las diferencias de sexo encontradas en este estudio son interesantes; sin embargo, es recomendable completar una investigación adicional para explorar esas relaciones.

Referencias

- Bar-Tal, D. y Bar-Zohar, Y. (1977), «The relationship between perception of locus of control and academic achievement», *Contemporary Educational Psychology*, **2**, 181–199.
- Beck, J. (1977), «Locus of control, task experiences and children's performance following failure», *Journal of Educational Psychology*, **71**, 207–210.
- Crandall, V. C. (1968), *Refinement of the IARQ scale*. (NIMH Progress Report, Grant No. MH-022 38, pp. 60–67).
- Crandall, V. C., Katkovsky, W., y Crandall, V. J. (1965), «Children's beliefs in their own control of reinforcement in intellectual-academic achievement situations», *Child Development*, **36**, 91–109.
- Diener, C. y Dweck, C. (1978), «An analysis of learned helplessness», *Journal of Personality and Social Psychology*, **36**, 451–462.
- Dweck, C., y Licht, B. (1980), «Learned helplessness and intellectual achievement», en J. Garber y M. E. P. Seligman (Eds.), *Human helplessness: Theory and applications* (pp. 197–221). New York: Academic Press.
- Dweck, C., y Reppucci, N. (1973), «Learned helplessness and reinforcement responsibility in children», *Journal of Personality and Social Psychology*, **25**, 109–116.
- Lefcourt, H. M. (1980), «Personality and locus of control», en J. Garber y M. E. P. Seligman (Eds.), *Human helplessness: Theory and applications* (pp. 245–259). New York: Academic Press.
- Lochel, E. (1983), «Sex differences in achievement motivation», en J. Jaspars, F. D. Finchman, y M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental and social dimensions* (pp. 193–220). London: Academic Press.
- McGhee, M. G., y Crandall, V. C. (1968), «Beliefs in internal-external control of reinforcements and academic performance», *Child Development*, **39**, 91–102.
- Nicholls, J. (1975), «Causal attributions and other achievement-related cognitions», *Journal of Personality and Social Psychology*, **31**, 379–389.
- Nowicki, S., y Strickland, B. (1973), «A locus of control scale for children», *Journal of Consulting and Clinical Psychology*, **40**, 148–154.
- Parsons, J. (1978), *Cognitive mediation of the effects of evaluative feedback on children's affect and expectancy for success*. Symposium paper presented at the annual meeting of the American Educational Research Association, Toronto.
- Parsons, J. C. (1981), *Attributions, learned helplessness and sex differences in achievement*. Paper presented at the annual meeting for the American Educational Research Association, Los Angeles.
- Rotter, J. B. (1966), «Generalized expectancies for internal locus of control of reinforcement», *Psychological Monographs*, **80** (1, Whole No. 60).

Source: Boss, M. W., y Taylor, M. C. (1989), «The relationship between locus of control and academic level and sex of secondary students», *Contemporary Educational Psychology*, **14**, 315–322.

Revisión de la bibliografía. Después de que los investigadores formulan un tema de investigación, realizan una búsqueda sobre estudios relacionados con el tema. La revisión resume y analiza la investigación previa y muestra cómo el estudio actual se relaciona con la bibliografía. La longitud de la revisión puede variar, pero debería ser selectiva y concentrarse en el modo en que el estudio actual contribuirá al conocimiento ya existente. Como en el ejemplo, debería ser lo suficientemente largo para demostrarle al lector que el investigador posee un conocimiento razonable de la relación entre lo que se ha hecho y lo que se hará. Generalmente no hay un encabezado separado que identifique la revisión, pero siempre está situado antes de la sección de métodos.

Hipótesis o preguntas de investigación. Tras la revisión bibliográfica, los investigadores sitúan la hipótesis o pregunta/s. Basándose en la información resultante de

la revisión, los investigadores enuncian una hipótesis que indica sus predicciones de lo que ocurrirá en el estudio. Las hipótesis pueden ponerse a prueba empíricamente y proporcionan enfoques para la investigación. Para algunas investigaciones no es adecuado hacer predicciones y en algunos estudios se formula una pregunta de investigación en lugar de una hipótesis. Tanto si es una pregunta como una hipótesis, el enunciado debería contener términos definidos objetivamente y establecer relaciones de una manera clara y precisa, tal como la hipótesis en nuestro ejemplo (las dos últimas líneas antes de la sección de metodología).

Metodología. En la sección de método o metodología, el investigador indica el diseño de la investigación, los sujetos, las pruebas y los procedimientos manejados en el estudio. Teóricamente, esta sección contiene suficiente información para permitir que otros investigadores rebatan el estudio. Normalmente hay un subtítulo para cada apartado de la sección de métodos.

En el apartado *sujetos* (algunas veces denominado participantes de la muestra o fuente de los datos), el investigador describe las características de los individuos de los que se ha recogido la información. Hay una indicación del número de sujetos y de la forma en que fueron seleccionados para el estudio.

El apartado de *pruebas* describe las técnicas que se han manejado para recoger la información. Debería ser una indicación de la validez y la fiabilidad de los resultados para cada instrumento de medida, que muestre que las técnicas son apropiadas para el estudio. Algunas veces, se incluyen ejemplos para ayudar al lector a comprender la naturaleza de la prueba.

El apartado de *procedimiento* o recogida de datos se utiliza para explicar cómo se ha llevado a cabo el estudio. Los autores describen cuándo se recogió la información, dónde y por quién. Describen qué se ha hecho a los sujetos y de que manera se recogió la información. Es importante proporcionar una descripción completa de los procedimientos. Es necesario que haya información suficiente para que el lector pueda saber cómo proceder para discutir el estudio. El procedimiento también afecta a la forma en que responden los sujetos. Por lo tanto, los lectores necesitan examinar cuidadosamente esta sección para interpretar los resultados.

Resultados. En la sección de *resultados* o *hallazgos* se presenta un resumen de los análisis de los datos recogidos. Esta sección puede parecer confusa para un investigador principiante debido al lenguaje estadístico, a los símbolos y convenciones que se emplean para presentar los resultados. Normalmente, se indican en tablas y gráficos dentro del texto del artículo. Los resultados deberían ser presentados objetivamente sin ningún tipo de interpretación o discusión, resumiendo lo que se ha encontrado. (Algunas veces, la interpretación irá a continuación de los resultados en esta sección). Puesto que la sección de resultados contiene información crucial en el artículo, el lector debe ser capaz de entender y evaluar el material. Esto es importante para evitar una aceptación indiscriminada de las conclusiones. En este momento no es necesario comprender la estadística presentada en nuestro ejemplo.

Exposición, implicaciones y conclusiones. Los investigadores indican en esta sección cómo se relacionan los resultados con el problema o la hipótesis de investigación. Es una interpretación no técnica de si los resultados confirman la hipótesis o responden a la pregunta de investigación. Si el estudio es exploratorio o contiene hallazgos no esperados, los investigadores explican por qué creen que han obtenido esos resultados. La explicación debería incluir un análisis de cualquiera de las deficiencias en la metodología empleada y una indicación a otra investigación que pueda explicar el porqué se obtuvieron ciertos resultados. Esta sección también se utiliza para indicar las implicaciones del estudio para la investigación futura y sus aplicaciones prácticas, y para proporcionar conclusiones globales. Esta sección se identifica por varias etiquetas diferentes, las más frecuentes son: *exposición, conclusión o resumen*.

Referencias. Las referencias y las notas de referencias que se citan en el artículo van a continuación de la exposición. El estilo puede variar. La revista en la que se publica nuestro artículo modelo usa el formato APA (*American Psychological Association*) de 1974.

Guías para la evaluación de la investigación cuantitativa

No hay ningún método o técnica consensuados para la lectura de artículos de investigación. Algunos lectores comienzan por las conclusiones y otros siguen la secuencia descrita. Nuestra experiencia sugiere que el lector debería empezar por el resumen, después la introducción, el tema de investigación y la conclusión. Si después de leer estas secciones, el lector todavía está interesado en el artículo, debería empezar por el principio y leer el artículo completo más cuidadosamente. Siempre que se lee una investigación, se debe tener en mente la importancia significativa o práctica del estudio. La investigación es significativa si no hay debilidades serias en el diseño y las diferencias obtenidas entre grupos o individuos o las relaciones apuntadas son lo suficientemente amplias como para sugerir cambios en la teoría o en la práctica.

Además, deberíamos tener en cuenta otras cuestiones en la lectura de investigación. Aunque en este momento necesite ponerse al corriente de estas consideraciones, sólo se puede llegar a una comprensión completa de las cuestiones y sus aplicaciones después de un estudio completo y detallado de cada tema. Las siguientes preguntas, organizadas de acuerdo a cada sección principal de un artículo de investigación, constituyen una guía para la evaluación de investigaciones cuantitativas³.

³ Adaptado de *Educational Research: Readings in Focus*, 2ª edición, de Irvin J. Lehmann y William A. Mehrens. Copyright © 1979 por Holt, Rinehart y Winston, Inc., reeditado con licencia del editor.

Tema de investigación

1. ¿Está enunciado el tema de forma clara y breve?
2. ¿Está suficientemente delimitado para facilitar su investigación? Al mismo tiempo, ¿posee suficiente valor teórico o práctico para justificar el estudio?
3. Posiblemente con la excepción de alguna investigación descriptiva, ¿se enuncia de tal manera que expresa la relación de dos o más variables?
4. ¿Está fundamentado? ¿El tema se ha estudiado con anterioridad? Si es así, ¿debería estudiarse, de nuevo, este tema? ¿Es probable que el estudio proporcione conocimientos adicionales?
5. ¿Darán lugar los resultados a hipótesis adicionales y, por lo tanto, incrementen la probabilidad de añadirse al conocimiento existente?

Revisión de la bibliografía

1. ¿La bibliografía ha sido revisada adecuadamente?
2. ¿La revisión presenta el material pertinente o sólo es relleno?
3. ¿La revisión evalúa críticamente los estudios y hallazgos previos o es sólo un resumen de lo que se conoce, sin apuntar ningún tipo de posibles deficiencias o explicaciones alternativas?
4. ¿Apoya la revisión la necesidad de estudiar este tema?
5. ¿Establece la revisión un marco teórico para este tema?
6. ¿Relaciona la revisión los estudios previos con el tema de investigación?

Hipótesis o preguntas

1. ¿Se avanzan algunas suposiciones con respecto a las preguntas o hipótesis? Si es así, ¿son explícitas (deberían serlo) o están implícitas?
2. ¿Las hipótesis son consistentes con la teoría y los hechos conocidos? ¿Son comprobables? ¿Proporcionan una respuesta al problema?
3. ¿Están todos los términos definidos adecuadamente en términos operativos?

Metodología

1. ¿El diseño, los procedimientos y las pruebas empleados para la recogida de datos están descritos con la suficiente claridad como para permitir a otro investigador objetar el estudio?
2. ¿Está descrita la población al completo? ¿El investigador ha utilizado la población total o sólo una muestra? Si utilizó una muestra, ¿es representativa de la población de la que fue extraída?
3. ¿Se presenta información sobre índices de validez y fiabilidad?
4. ¿Se ha empleado un pretest? ¿Se realizó un estudio piloto? Si es así, ¿por qué? ¿cuáles fueron los resultados? ¿Se modificó el tema, la cuestión o el procedimiento como resultado del pretest o del estudio piloto y, si fue así, era justificable o deseable?
5. ¿Hay alguna debilidad obvia en el diseño del estudio?

Resultados

1. ¿Era necesario utilizar técnicas estadísticas para analizar los datos? Si es así, ¿las técnicas empleadas eran las más apropiadas?
2. ¿Se presentaron los datos de forma adecuada?

Exposición, implicaciones, conclusiones

1. ¿Las generalizaciones y conclusiones son compatibles con los resultados? ¿cuáles son sus implicaciones? ¿El investigador ha generalizado en exceso sus resultados?
2. ¿Describe el investigador las limitaciones del estudio?
3. ¿Hay variables accidentales que puedan haber afectado a los resultados? ¿Han sido tenidas en cuenta por los investigadores?
4. ¿Las conclusiones presentadas son compatibles con la teoría o los hechos conocidos?
5. ¿Se han presentado adecuadamente y discutido las conclusiones (tanto las relevantes para la hipótesis original como los hallazgos fortuitos)?

Cómo leer investigación cualitativa: un ejemplo etnográfico

Existe más diversidad en los formatos empleados para exponer una investigación cualitativa que en los de los estudios cuantitativos. Esto es así debido a los muchos tipos de estudios cualitativos y al hecho de que, hasta hace poco, no se realizaban informes de investigación cualitativa educativa. Aunque no hay una forma única de presentar una investigación cualitativa, muchos de los informes publicados tienen las secciones principales que describimos a continuación. Sin embargo, en contraposición con los estudios cuantitativos, estas secciones pueden no ser fácilmente identificables o bien ser identificadas por términos descriptivos relacionados con el tema. Ver extracto 2.2 para un ejemplo de informe de investigación cualitativa.

Introducción. La introducción presenta el marco general del estudio, indicando la importancia potencial de la investigación. Resume las intenciones generales del investigador, junto a un enunciado general del asunto o propósito de la investigación. Para un artículo de revista, normalmente, sólo se señala uno de los múltiples enfoques posibles. La introducción incluye una revisión preliminar de la bibliografía para presentar posibles marcos teóricos que serán útiles para entender los datos y los resultados. La revisión justifica la necesidad de un estudio de caso descriptivo. La introducción también puede indicar la estructura del resto del informe.

Metodología. La sección de metodología describe el diseño del estudio, incluyendo la selección y descripción del escenario, la función del investigador, la entrada inicial para la observación, la duración del estudio, el número de

participantes y cómo fueron seleccionados y las estrategias de recogida y análisis de datos. Esta información es necesaria para evaluar la solidez de los procedimientos. La cantidad de detalles contenidos en esta sección puede variar dependiendo del tipo de informe de investigación. En una cantidad relativamente pequeña de artículos publicados la metodología puede formar parte de la introducción.

Resultados e interpretaciones. En esta sección, el investigador presenta los datos que fueron recogidos; normalmente lo hace en forma de largas narraciones y analiza los datos. Esto debería hacerse con el suficiente detalle como para permitir al lector juzgar la exactitud del análisis. Los datos se usan para ilustrar y consolidar las interpretaciones del investigador. El análisis se entremezcla, a menudo, con la presentación de los datos. Los datos a menudo aparecen en forma de citas de los participantes. Es importante indicar el propósito del análisis de datos y describir lo que se ha aprendido sintetizando la información. Debido a que las presentaciones están en forma de narración, frecuentemente, se incluyen subtítulos descriptivos que connotan los diferentes hallazgos.

EXTRACTO 2.2

EJEMPLO DE UN INFORME DE INVESTIGACIÓN CUALITATIVA

¿PUEDE SER UNA CARRERA PROFESIONAL LA ENSEÑANZA EN EDUCACIÓN PRIMARIA? UNA BÚSQUDA DE NUEVAS FORMAS DE ENTENDER EL TRABAJO DE LAS MUJERES

Sari Knopp Biklen, *Syracuse University*

INTRODUCCIÓN

¿Es apropiada la noción predominante de carrera profesional en nuestros días? Para examinar esta cuestión, veamos primero cómo la bibliografía y la sociedad establecen el concepto de carrera. Comparemos entonces esta perspectiva con la de un grupo particular, las mujeres que enseñan en la escuela primaria. A través de las biografías de las informantes, vemos cómo el concepto tradicional resulta inadecuado para organizar las vidas de estas mujeres. Finalmente, sugerimos qué elementos debe incluir un concepto revisado de carrera profesional¹.

El trabajo actual, tanto en educación como en las cuotas por sexo, hace del momento presente el ideal para examinar estas inquietudes. El interés tanto académico como federal por la calidad de la educación ha dirigido la atención hacia las escuelas y, particularmente, hacia los maestros². Sin embargo, se ha prestado escasa atención al modo en que las diferencias de género se relacionan con los intereses educativos³. Esta omisión es desconcertante ya que la enseñanza primaria es considerada una ocupación para mujeres.

Los analistas demográficos sugieren que las oportunidades de desarrollo laboral para las mujeres han contribuido al abandono de este campo por parte de mujeres

brillantes⁴. La sociedad se pregunta si la falta de solicitudes para enseñar de personas bien preparadas no estará creando un cuerpo docente mediocre. Conjuntamente, existe el desprecio público para aquéllos que quieren pasar el día con niños pequeños. Por consiguiente, los políticos tratan de descubrir cómo atraer hacia la enseñanza a personas brillantes y creativas, ciertamente, un problema importante.

Carreras: definiciones tradicionales

A pesar de los cambios en la mano de obra, de la apertura de los campos que antes se resistían más a las mujeres y de la incorporación de las mujeres a puestos profesionales y de gestión más elevados, la estructura de la carrera está basada en las formas en las que los hombres han podido vivir sus vidas, libres de la responsabilidad hacia la familia. Uno empieza a trabajar, después de la universidad o la escuela de postgrado o de una formación médica o científica avanzada y trabaja continuamente, ascendiendo en fases graduales. Desde esta perspectiva, una carrera es «un patrón preestablecido de actividad profesional organizada, con un movimiento hacia arriba, través de fases preparatorias conocidas y un avance basado en los méritos y honores»⁵. Este punto de vista resalta la coherencia de la carrera y la orientación hacia objetivos del trabajador⁶. Podemos estudiar, por lo tanto, la estructura de carreras particulares como la médica, la jurídica y la docente⁷.

Intervienen dos elementos fundamentales. El primero es la promoción⁸. Esta perspectiva presenta las carreras como trayectorias medibles en la esfera pública. La carrera profesional está definida por la naturaleza de la participación en el sistema de salario y por el estatus que se deriva de esta participación⁹. Visto de esta manera, la enseñanza, por lo menos, es «semiprofesional» en el mejor de los casos, si no marginal¹⁰.

El segundo elemento es el compromiso laboral. Se entiende por compromiso la dedicación al trabajo. Se supone que la vida personal o familiar debe ajustarse a las demandas laborales. El compromiso tiene dos aspectos principales, ambos relacionados con el tiempo. El primer aspecto es el compromiso a largo plazo. Aquí, el compromiso laboral se refiere a la decisión de situar la carrera en primer lugar. Por lo tanto, si una mujer decide llegar a ser profesora en lugar de directora, se diría que ha disminuido su compromiso¹¹. Asociado a este aspecto del compromiso laboral está la cantidad de tiempo que la mujer espera trabajar. Por ejemplo, Mason midió el compromiso laboral de los profesores preguntándoles dónde planeaban estar dentro de cinco años¹².

El segundo aspecto del compromiso laboral se relaciona con el uso a corto plazo del tiempo de trabajo. Un baremo común para diferenciar las ocupaciones de las carreras profesionales es, por ejemplo, la tendencia de las carreras a sobrepasar la franja horaria de 9 a 17. Un ejemplo de esto son los doctores:

La habilidad de un campo laboral para inducir a sus miembros a hacer su trabajo sin importar el tiempo que conlleve ni lo que otros les pidan es fundamental para el estatus de una profesión. Estas exigencias son rigurosas, por ejemplo, en la medicina, donde las vidas pueden estar en peligro por la decisión de un médico acerca de la prioridad para terminar un trabajo¹³.

Nuestra noción de carrera profesional incluye el trabajo firme y el compromiso a lo largo de un extenso periodo de tiempo¹⁴. Las mujeres deben aceptar este punto de vista en las carreras profesionales si quieren tener éxito: «En la medida en que los valores del trabajo masculino son dominantes, las mujeres que desean tener éxito deben aceptar las definiciones y expectativas masculinas sobre los compromisos laborales»¹⁵. Las implicaciones para las mujeres son directas bajo este punto de vista. Para las mujeres, situar el compromiso laboral por encima de la familia para, digamos, dar prioridad directamente a sus exigencias laborales significa actuar contra las normas sociales¹⁶.

La promoción y el compromiso laboral, por lo tanto, son dos elementos esenciales de la comprensión externa de la carrera¹⁷. Examinaremos las perspectivas de trabajo de un grupo de profesores de primaria para describir un punto de vista alternativo. Sin embargo, primero vamos a examinar las vidas laborales de los profesores.

Profesores y carreras

Se ha descrito la enseñanza como la «falta de carrera»¹⁸ y como una ocupación de «carreras laterales»¹⁹. El cambio de colegio puede significar un cambio de carrera lateral. Los maestros, por ejemplo, pueden considerar un traslado a un colegio en un vecindario de mayor nivel socioeconómico como una promoción. La etiqueta de «falta de carrera» se refiere a la estructura de la enseñanza; a saber, la imposibilidad de promoción dentro de la enseñanza. Un avance significa llegar a ser administrador o director y abandonar la enseñanza. Además, la enseñanza está estructurada para acomodar los patrones «dentro y fuera» del empleo de las mujeres, esto es, «persistir en la enseñanza es, en cierto sentido, “relegar” una posición superior o el matrimonio»²⁰. Desde este punto de vista, un compromiso de por vida hacia la enseñanza evidencia un fracaso en lugar de un éxito.

Se ha advertido ampliamente el bajo compromiso laboral de los maestros de primaria y se ha relacionado este hecho con el género²¹. Esta bibliografía refleja la noción de las expectativas basadas en las normas sobre el papel femenino dentro del entorno ocupacional. La enseñanza se convierte en una prolongación del papel de la mujer.

La investigación más reciente sobre mujeres y trabajo, abordada desde la nueva perspectiva de estudios sobre la mujer, ha cuestionado este enfoque²², sugiriendo que las formas en las que hemos examinado las vidas laborales femeninas resultan inadecuadas y engañosas debido a que están basadas en suposiciones estereotipadas sobre las mujeres²³. Desde este punto de vista y debido a nuestra inmersión en una sociología de las ocupaciones que toma las vidas de los hombres como norma, no podemos pensar claramente en el trabajo que realizan las mujeres²⁴.

Los estudiosos que trabajan desde esta perspectiva han observado el tema del compromiso laboral de forma diferente. Han examinado factores en el lugar de trabajo que pueden determinar experiencias diferentes para las mujeres²⁵. Han argumentado que el poder influye en el comportamiento más que el género²⁶ y han situado el compromiso laboral en un marco diferente²⁷. Han criticado los modelos que los sociólogos han trasladado al estudio del trabajo de hombres y mujeres²⁸. Algunos han analizado las insinuaciones de los medios de comunicación a las mujeres trabajadoras²⁹. Puede verse este estudio en el contexto de las nuevas investigaciones sobre las mujeres.

METODOLOGÍA

Métodos y participantes

Los temas que se discuten en este artículo forman parte de un estudio más amplio centrado en cómo consideran su trabajo un grupo de maestros de la escuela primaria, qué valoran o critican de su trabajo y cómo compatibilizan sus intereses laborales con las expectativas familiares y el papel sexual³⁰. El investigador confía en los métodos cualitativos de observación participante y la entrevista en profundidad³¹. La naturaleza inductiva de esta aproximación metodológica y el énfasis en la perspectiva del participante hace que el investigador busque las maneras en las que las personas estudiadas dan significado a sus experiencias. Los investigadores cualitativos asumen que las personas actúan sobre la base de sus interpretaciones de la experiencia. Por lo tanto, están interesados en qué experimentan los sujetos y en cómo interpretan estas experiencias³².

Las entrevistas y observaciones se llevaron a cabo en escuelas durante un periodo de ocho meses en un entorno principal y uno auxiliar. Ambas escuelas estaban localizadas en una ciudad de tamaño medio del noreste de Estados Unidos. El entorno principal, Escuela de Primaria de Vista City³³, tenía una población estudiantil de 800 alumnos y una reputación académica sobresaliente. Cuando el periódico local publicaba, cada primavera, los resultados de los tests de rendimiento, la Escuela de Primaria de Vista City siempre alardeaba de las puntuaciones más altas. Los datos publicados en este artículo proceden de este entorno principal.

Todos los maestros de Vista City eran mujeres menos dos. El rango de edad estaba entre la veintena y el final de los cincuenta. Había solteros, casados y divorciados. La mayoría de las mujeres solteras habían trabajado sólo en la enseñanza, aunque dos de ellas habían intentado realizar otro tipo de trabajos durante un corto periodo de tiempo con la esperanza de un mayor reconocimiento. Aquéllos que estaban casados, tenían familias y niños que eran importantes en sus vidas. Igualmente, entre las mujeres casadas había variedad en sus patrones de actuación a largo plazo. Algunas de las profesoras casadas tenían maridos que deseaban familias tradicionales y tenían valores conservadores sobre el papel de las mujeres. Algunas de las profesoras de más edad tenían padres conservadores que les prohibieron entrar en el mercado laboral cuando eran jóvenes. Incluso, otras sólo se habían tomado seis semanas de baja por maternidad o habían permanecido en casa unos pocos años antes de volver al trabajo. En un caso, una profesora no terminó sus estudios universitarios hasta que sus hijos fueron a la escuela.

También había un abanico de intereses y participación en los problemas de mujeres. Dos de las profesoras participaban activamente en la Organización Nacional para las Mujeres. Otras manifestaban cierto interés en factores que facilitan el trabajo de las mujeres o que promueven el ascenso de las mujeres. Ninguna de las solteras entrevistadas mostraba interés alguno en los temas relacionados con las mujeres; aunque algunas de las casadas sí lo hacían, asumían la responsabilidad principal tanto en sus casas como en sus clases. No podían ser descritas como rebeldes contra las normas sociales.

Caracterizando sus vidas como «carreras interrumpidas» y por deferencia a los valores sociales de sus maridos, es difícil entender cómo se puede argumentar que estas

mujeres entienden su trabajo como una carrera docente. A pesar de todo, sugeriré que la forma tradicional de evaluar una carrera o un compromiso laboral (como un movimiento continuo ascendente a través de distintas fases ocupacionales) no hace justicia al punto de vista de estas mujeres. En la próxima sección, donde examinamos cómo los maestros describieron su trabajo, nos situamos en la estructura interna de la carrera.

Los maestros describen su trabajo

Durante los primeros meses de entrevistas y observación, cuando se les preguntaba a los maestros cómo describirían sus experiencias en el trabajo, a menudo decían: «siempre he trabajado». Durante las primeras fases de mi investigación, me formé una imagen mental de una persona que había trabajado siempre a tiempo completo con una meta profesional muy clara. Me recostaba en mi asiento esperando escuchar los detalles y pensando para mí misma: «estoy cerca de encontrar algo. Esto ciertamente contradice lo que he leído acerca de los patrones de trabajo de las mujeres». Sin embargo, cuando las mujeres me describían sus historias laborales, casi siempre hablaban de un patrón discontinuo de crianza de los hijos, trabajo a tiempo parcial y, finalmente, reincorporación a tiempo completo al mercado laboral. Lo más sorprendente es que siempre terminaban sus historias reiterando que siempre habían trabajado.

Kate Bridges era una de las profesoras que se describió a sí misma como trabajando ininterrumpidamente; pero, que detalló su historia laboral real con este patrón discontinuo. Describió sus sentimientos sobre su trabajo: «He sentido pasión por la enseñanza durante 20 años». Puede que hubiera sentido pasión; pero, exactamente, no había estado enseñando durante 20 años.

Ella recordó «enseñar de forma constante desde 1960-61», pero tuvo bajas por maternidad. «Volví a casa por un año y me dije: ‘madre en casa para siempre’. Y entonces, de repente, volví a la enseñanza de nuevo». Ha realizado varias tareas docentes:

Empecé en los colegios públicos de San Francisco, enseñé allí durante dos años en una especie de colegio del barrio pobre y lo adoraba. Cuando llegaron los hijos busqué un trabajo a media jornada y trabajé con niños con dificultades. Parecía ideal ya que tenía una amiga con niños de la misma edad y ambas éramos profesoras. Decidimos montar una guardería y entonces ella se quedó embarazada de nuevo y me dejó sola. Yo lo hice durante dos años y me entusiasma. Era mi propia escuela.

Se trasladó a Chicago por el trabajo de su marido y consiguió un trabajo a media jornada en el departamento de educación de una universidad para enseñar a profesores. En este momento, había «obtenido casi la jornada completa y usted sabe que cualquier trabajo de media jornada es un horario de jornada y media en cierto modo». Se mudó por última vez hasta Vista City, de nuevo por el trabajo de su marido. Permaneció «principalmente» en casa durante siete años, aunque trabajó a tiempo parcial en un centro de rehabilitación de drogadictos desarrollando su programa educativo. Entonces sintió que debía volver a la enseñanza. Obtuvo un trabajo de profesora a jornada completa en la escuela primaria. Describía su determinación en aquella época: «voy a empezar mi carrera

docente y si no acepto ahora este trabajo, sabe Dios si volveré a tener esta oportunidad»³⁴. Como indica el comentario anterior, Kate reconocía formalmente que el comienzo de su carrera educativa fue cuando empezó a enseñar a tiempo completo en la escuela primaria. Sin embargo, sus conflictos surgían de la percepción de que había estado interesada, implicada y pensando en la enseñanza durante 20 años. Tal como ella lo veía, ella había permanecido comprometida con la enseñanza por su constante dedicación interna. Ella siempre había pensado en sí misma como una maestra.

Kate había trabajado mucho durante 20 años y alguno de estos trabajos implicaba la educación de niños en colegios. Pero, no siempre podía trasladar su pasión por la enseñanza a la realidad laboral. Aunque siempre se había considerado una maestra, no siempre estuvo contratada para enseñar a jornada completa. Y éste es el *quid* del asunto. Tal como lo entendían Kate y sus compañeras, cuidar de los niños propios, necesariamente, no se refleja en el propio compromiso laboral.

Hay que insistir en que estas profesoras no eran mujeres que contravenían las normas sociales, eran tradicionales. Sin embargo, al mismo tiempo, valoraban su identidad laboral como profesoras y no querían tener que elegir entre trabajo y familia. El perfil de carrera que Kate revela que está definido por la consistencia interna sobre una ocupación en vez del empleo externo permanente³⁵.

Profesoras y sus maridos

Los maridos influían esencialmente en la forma en la que las profesoras casadas tomaban decisiones sobre su trabajo. Algunas mujeres como Kate hablaban directamente con sus maridos sobre sus intenciones laborales. Otras mujeres tenían maridos que se resistían a los intereses laborales de sus esposas. Los casos de tres mujeres cuyos deseos de trabajar a jornada completa fuera de casa les obligó a desarrollar estrategias para superar la resistencia de sus maridos revelan las tácticas que desarrollaron para permitirles volver al mercado laboral. Sus situaciones son valiosas para el estudio porque, externamente, estas mujeres parecen las más complacientes con el papel de las mujeres.

Sylvia Richardson, a sus cincuenta años, había enseñado a los alumnos de séptimo curso durante cinco años antes de tomarse una excedencia de nueve años para criar a sus hijos, «y esperé hasta que mi hijo menor estuvo en el colegio». Desde entonces ha estado enseñando durante nueve años. Su marido no quería que volviese al trabajo porque le gustaba tenerla en casa. Ella dijo: «No estamos por esto de la liberación de las mujeres». Como su deseo por volver al aula se fortaleció, decidió que, para ganar flexibilidad, debería empezar por hacer sustituciones. En ese momento, una de sus viejas amigas maestras le dio un pequeño consejo: «Escucha, no hagas sustituciones en séptimo curso porque los alumnos son muy diferentes a cuando tú estabas allí». Entonces Sylvia decidió que, «si los alumnos de séptimo y octavo curso realmente eran violentos y si va a ser una situación diaria difícil para mí, mejor escojo un grupo que no tenga estas exigencias». Entonces su marido no sería capaz de decirle: «No necesitamos estos problemas. Vuelve a casa». El hecho que le facilitó cambiar las sustituciones, que su marido no consideraba como jornada completa, por la enseñanza a jornada completa, fue su disponibilidad. Como empezó a hacer sustituciones diariamente, finalmente, su marido cedió a que

volviese a la jornada completa. «Por lo menos sabremos dónde encontrarte», le dijo su marido.

Sylvia Richardson empleó una estrategia para volver a la enseñanza. Eligió un nivel académico menos «difícil» que le permitía dejarse los problemas laborales cuando volvía a casa. Sylvia representa un ejemplo interesante de cómo las mujeres pueden utilizar estrategias para cumplir sus metas cuando no quieren cambiar las normas sociales.

Mientras Sylvia empleaba cada oportunidad que tenía para desmarcarse del feminismo, al mismo tiempo, compartía las apreciaciones feministas del daño a la autoestima que podía proporcionar el ser ama de casa a jornada completa. Como señaló: «realmente es muy duro para cualquiera sólo estar todo en día en casa y hacer las tareas del hogar y cuidar a los niños porque te sientes fatal contigo misma. Pero, por otro lado, la familia necesita atenciones»³⁶. Aunque la estrategia elegida por Sylvia le había reportado beneficios, los costes que otras tuvieron que pagar fueron altos.

Tomemos el caso de Jessica Bonwit. Ella también dejó la enseñanza cuando nacieron sus tres hijos, aunque organizó un programa de guardería familiar durante este periodo. Pero, «algo hizo clic después de seis años» y supo que tenía que volver a dar clases en un colegio. Sin embargo, su marido, no compartía esta idea. Jessica le prometió que su vida no cambiaría y que cumpliría con sus compromisos de la casa. Como ella dijo: «A él no le importó demasiado que yo volviese a trabajar, porque su vida no cambió con respecto a antes de que yo volviera al trabajo, pero también sabía que yo tenía un interés real por mi trabajo y que era realmente feliz haciéndolo»³⁷. El precio que Jessica tuvo que pagar por trabajar a tiempo completo fue el de cargar con las responsabilidades tanto del colegio como de la casa. Estaba resentida por el egoísmo de su marido: «quiero a mi marido, pero a veces no me gusta demasiado». La enseñanza era tan importante para ella que, al menos hasta el momento, deseaba llevar esta carga. El impacto de su decisión sólo se verá en el futuro.

Carrie Amundsen no había trabajado nunca antes de tener hijos; pero, como las otras dos mujeres, también tuvo que desarrollar una estrategia para conseguir un empleo a jornada completa. Aunque Carrie estaba en los cincuenta, sólo había enseñado durante 10 años. Su marido nunca quiso que ella trabajara. De hecho, confesó que estuvo a punto de no casarse con él debido a su punto de vista sobre las mujeres trabajadoras. Ella se refería a él como «el cerdo macho chauvinista». Tuvo que encontrar una forma de facilitarse la entrada en el mercado laboral. Dijo, «así es como lo hice».

«Realmente, comenzó, ni siquiera estaba planeando volver al trabajo, nunca terminé mi licenciatura, así que asistí a un curso de Matemáticas en la universidad». Supuestamente no tenía ninguna meta en la cabeza excepto «enriquecimiento», pero «le encantó» el curso y cuando terminó realizó otros cursos más. Cuando aprendió todas las matemáticas posibles, «de algún modo, hizo algunas conexiones» entre las experiencias de sus propios hijos en el colegio con respecto al aprendizaje de la lectura y las dificultades especiales que tenían otros niños. Además, existían razones específicas para estas dificultades. Su interés se despertó. Cuando terminó su diplomatura obtuvo un master en Educación especial.

En ese momento, había adquirido experiencia en los colegios y quería ponerla en práctica. Un especialista en su área empezó un programa en la ciudad donde vivía y le

pidió que trabajara en él. Ella le dijo: mi marido nunca me permitirá trabajar a jornada completa». Su futuro patrón le respondió: «déjame que yo me ocupe de tu marido». Así, ella y el profesor de la facultad «hicieron un trato» en el que debería decirle a su marido que le gustaría intentarlo sólo durante un año. «Surtió efecto» y «le encantó» enseñar. Al finalizar el año ella «actuó como una comadreja» y continuó otro.

Estas tres profesoras eligieron tácticas para reconducir su trayectoria y enseñar sin tener que librar una batalla ideológica con sus maridos o rebelarse contra las normas sociales. Con sus propios esquemas temporales tradujeron su compromiso mental a una realidad laboral. Mientras estuvieron lejos de su trabajo de las instituciones escolares, las mujeres a menudo organizaron guarderías en casa, permanecieron en contacto con sus colegas y discutieron sobre la vida educativa, planearon volver al trabajo y, algunas veces, planearon estrategias para vencer la resistencia de sus maridos hacia su trabajo.

Externamente, las biografías de estas tres mujeres representan el patrón de carrera interrumpida. Sin embargo, internamente, ellas se consideraban profesoras estuviesen o no en el mercado laboral y tomaban decisiones que las mantenía en contacto con los niños o con los asuntos educativos. Aunque lo que hicieron no desafiaba, en cierta medida, los límites de sus vidas, demostraron una gran coherencia en su actitud hacia su trabajo.

Estructura externa y conceptos internos

Claramente, existe una manera diferente de opinar sobre la coherencia de la vida laboral de alguien. Estamos acostumbrados a pensar en la coherencia de la vida laboral en términos de promoción continua. Estas mujeres, por otro lado, manifestaron, a través de sus vidas, que querían tener hijos y familia, así como ser profesoras. Se cuestionaban por qué las bajas maternales para cuidar a sus hijos se equiparan a la pérdida de su compromiso laboral cuando siempre se habían sentido profesoras. La estructura externa entra en conflicto con sus pensamientos internos.

Sin embargo, como hemos dicho, el compromiso laboral, en el concepto tradicional de carrera, es inseparable de la promoción. Pero, de nuevo, muchas de las profesoras observadas tienen una perspectiva diferente de estos asuntos. Como hemos visto, las profesoras se centraban en los niños y este interés era una parte determinante de su orientación laboral.

Los profesores más comprometidos de la Escuela de Primaria de Vista City tenían un nivel alto de idealismo hacia su trabajo. Estas expectativas les hacían trabajar duro para conseguir sus metas, además de contribuir a su reputación como profesores excelentes. Al mismo tiempo, el entorno laboral provocaba frustraciones porque no coincidía con sus concepciones de cómo querían trabajar. Algunos de estos profesores se concentraban en su trabajo dentro del aula, aislándose de otros adultos en el edificio. No todos los profesores reflejaban ese patrón, pues alguno trasladaba su idealismo a otros aspectos del trabajo, como la tarea del comité. Sin embargo, en ningún caso el compromiso alto de los profesores o su idealismo impulsó sus carreras. Existían pocas oportunidades de promoción disponibles para lo que era la estructura explícita de la profesión. El idealismo dio lugar a que los profesores se concentraran en la calidad del trabajo y en el ambiente laboral.

Ya que las oportunidades administrativas son escasas, pocos profesores afrontan la situación de decidir qué hacer cuando se presenta la oportunidad. Sin embargo, aquéllos que se han enfrentado a tal decisión muestran el idealismo sobre la enseñanza que otros profesores, a menudo, exhiben en su trabajo diario. Por ejemplo, cuando a Barbara Timmitts se le ofreció el puesto de especialista educativo, se vino abajo. Sentía que no había enseñado lo suficiente para hacer bien el trabajo. Cuando se le ofreció la segunda vez, lo aceptó, no porque estuviese segura que lo iba a hacer excelentemente, sino porque quería el trabajo y tenía miedo de que no se lo ofrecieran una tercera vez.

Si examinamos lo que piensa Barbara Timmitts sobre su oferta de trabajo, nos daremos cuenta de que su mayor preocupación era si podría o no hacerlo bien, en lugar de pensar que el puesto de trabajo podría servirle como avance en su carrera. Este tipo de orientación laboral puede definirse como idealista porque refleja el concepto ideal de una persona sobre cómo debe hacerse un trabajo. Los profesores se centraron en el contenido de la profesión, no en si su trabajo era un acceso a otras opciones profesionales. La calidad de ejecución está por encima del valor de la carrera. Estos profesores pensaban a menudo en cómo podían servir a su profesión en lugar de cómo su profesión podría servirles a ellos.

Christine Bart no tenía un puesto administrativo. Presidió el equipo de primer curso, sin embargo, no recibió ninguna remuneración extra por ese trabajo. Sin embargo, al igual que Barbara, desempeñó un papel importante entre los profesores, tanto del colegio como de la junta de distrito. Al igual que Barbara, su idealismo dio forma a los compromisos que emprendió. Cuando le pregunté a Christine sobre sus metas de liderazgo, dijo que no estaba interesada en llegar a ser directora. Desde su punto de vista, los directores parecían no tener poder. Entre las demandas de los padres y las regulaciones de la administración central, tenían las manos atadas.

Bart no estaba segura de que lo que hacen los gestores sea efectivo. Ella, al igual que muchos profesores de Vista City, compartía el punto de vista de que, nadie puede ser productivo o eficaz como administrador o gestor, y que esos puestos derrochan un talento valioso. Veía a Barbara Timmitts como el caso en cuestión: «mi prioridad está aquí, con estos niños. Miras a alguien como Barbara Timmitts. Es la mejor profesora de este colegio y le pusieron a empujar un carro por los pasillos. Pasea con formularios de petición y un lápiz en su mano. ¿Y, ahora qué pasa?». Los niños son el centro del trabajo.

Amelia Dickenson había dirigido el programa de superdotados antes de que la trasladaran de vuelta a las clases. Cuando le pregunté si le había gustado volver a dar clase a tiempo completo, me contestó: «es realmente maravilloso; ¡me entusiasma! Realmente te sientes como si estuvieses llevando algo a cabo»³⁸. El compromiso de los profesores con su trabajo y su intento por hacer su trabajo de acuerdo con su ideal de cómo debe hacerse, también se manifestaba en el trabajo cotidiano. Aunque los profesores, a menudo, se quejaban de los pocos momentos de descanso que tenían durante el día, muchos empleaban estos momentos para dedicarlos a la enseñanza. Por ejemplo, Roberta Blake, decía que no tenía tiempo libre durante el día, ni siquiera durante la comida. Explicó que creía en la integración, así, en lugar de hacer que los alumnos de su clase de educación especial fueran al grupo dedicado a «los especiales» (arte, música o gimnasia), enviaba a unos pocos cada vez a diferentes clases ordinarias. Por lo tanto, sus ideas le

generaban más trabajo. Como ella señaló: «después de todo, estos alumnos están aquí para eso (para ser integrados)».

Era raro que Kate Bridges fuera a la sala de profesores para comer. No era sólo porque le molestara el humo. Decidió, además, que devolver por la tarde los trabajos que los alumnos habían completado por la mañana mejoraba los resultados académicos. «Tengo tantos trabajos que voy acumulando durante la mañana. Si puedes devolverlos para que los niños trabajen en ellos tan pronto regresan de la comida, ¿por qué no corregirlos y dar un *feedback* inmediato? Como digo yo, el kilometraje es mucho mayor con este sistema. Sin embargo, te dejas la piel. Por eso aquí estoy sentada comiéndome un bocadillo y corrigiendo esos trabajos, además, me doy cuenta que es una educación mejor (y no tendré que corregir todos estos trabajos por la noche)». Kate veía su elección como un «trueque»; sin embargo, quedarse en su clase la aislaba de las relaciones escolares.

El equipo de sexto curso también renunció a poder tener un descanso durante el día. Le pidieron a la dirección que organizara a todos sus alumnos «especiales» a las 8:15 (el colegio empezaba a las 8:00), así podrían pasar el resto del día con sus alumnos. Ya que éstos «cambiaban» de clases para las diferentes materias, los miembros del equipo dijeron que no tenían «suficiente tiempo con ellos». Aparentemente querían el máximo tiempo con sus alumnos.

Sandra Miller renunció a la pausa de la comida durante un mes, para hacer sus «evaluaciones» ya que quería realizar las pruebas a sus alumnos de la manera que pensaba debían hacerse, «cuando no hay nadie alrededor». De esta manera, ella tampoco tendría que reducir el tiempo de sus grupos de lectura para administrar las pruebas.

En la Escuela de Primaria de Vista City, estos «descansos» no sumaban mucho tiempo. Los profesores tenían tres periodos «libres» a la semana mientras sus alumnos asistían a clase de arte, de música y de gimnasia. Desde la perspectiva de estos maestros, el día no estaba estructurado para maximizar la educación de sus alumnos. Los profesores confesaban que sentían que alcanzaban sus metas sólo cuando sacrificaban su propio tiempo. Este sacrificio no significaba que no quisieran los descansos, pero poniendo las cosas en una balanza, preferían crear condiciones óptimas de enseñanza.

Sacrificar su tiempo no es la única manera mediante la que los profesores intentan llevar a cabo su ideal de cómo debe ser una buena enseñanza. Su idealismo se expresaba de muchas otras maneras. Una profesora ejerció de tutora durante el verano anterior porque quería adquirir esa experiencia particular, incluso, aunque estuviese mal pagado. Resumió sus vivencias con «y, realmente aprendí mucho». Jessica Bonwit rechazó un trabajo educativo cuando regresó a enseñar, después de un tiempo en casa con sus hijos, debido a que consideraba que no poseía los requisitos mínimos para una buena situación de enseñanza. Comentó que no significaba que le resultara difícil sino que deseaba aproximarse a lo que ella consideraba una buena situación educativa.

Una profesora ayudante abandonó el puesto de trabajo estable que consiguió porque el niño especial con el que trabajaba fue trasladado a otro distrito. No quería que fuese trasladado porque estaba a mitad de curso. Sus intentos por posponer el traslado hasta el año siguiente fueron inútiles. Podría haberse quedado en el distrito mientras el niño

tenía otro maestro en el nuevo distrito, pero no quería poner en peligro todos los progresos que había realizado el niño ese año. Para ella lo más importante era terminar el trabajo que había empezado. Aunque la escuela a la que fuera no tuviese un programa para el niño, estaba decidida a continuar haciendo que la integración funcionara con el niño: «vamos a hacer este trabajo. Incluso si no tiene un programa allí, haremos que sea un éxito. Cueste lo que cueste, haremos un buen programa para Jacob allí».

Estos profesores centraban su esfuerzo en el contenido del trabajo, no en conseguir un ascenso. Por lo tanto, sus mayores frustraciones no venían de que se desvanecieran sus esperanzas de ascenso, sino más bien de que les forzaran a tomar decisiones que sentían que ponían en peligro su visión educativa³⁹.

Los profesores del Vista City trasladaban sus ideales ocupacionales a su lugar de trabajo. Estos ideales estaban centrados en el contenido del trabajo más que en sus potenciales escalafones laborales. Quizás, su idealismo se refleja en el lenguaje que una profesora empleó para describir su primera experiencia educativa: «fue un “matrimonio” entre ella y los niños; ésta es la clase de relación que buscaban muchos profesores».

El idealismo de los profesores sobre su trabajo también afectaba a sus aspiraciones sobre la importancia que tendría ser director. Los profesores del Vista City eran particularmente críticos con los de las oficinas centrales, ya fueran los empleados que planificaban los cursos de perfeccionamiento del personal o de la administración de educación especial, que eran criticados por estar desconectados del personal de la escuela y ser incapaces de hacer frente a problemas difíciles. Desde la perspectiva de estos profesores, con frecuencia, los que llegaban a ser directores y empleados de las oficinas centrales eran profesores mediocres. Ya que los problemas y demandas también eran grandes allí, no veían razones de peso para anhelar esos puestos de trabajo.

Sus metas eran llegar a ser lo que llamaban «grandes profesores». Un gran profesor tenía buena fama, en su escuela o en su distrito, por sus méritos. El estilo no determinaba la fama de un profesor. Algunos eran más estrictos que otros, o estaban más o menos interesados en, digamos, centros de aprendizaje. Eran profesores en cuyas clases los padres deseaban que estuviesen sus hijos, porque en estas clases la mayoría de los niños aprendían, eran estimulados y felices.

Aquéllos con reputación de grandes profesores eran capaces de asumir más poder y afianzar su autonomía. Eso suponía una gran ventaja. Por ejemplo, formalmente eran elegidos para el claustro escolar y las comisiones del distrito. De manera informal, el director solicitaba sus puntos de vista y les daba carta blanca para organizar sus clases y su currículo. Pero lo principal era la reputación y la satisfacción personal que generaban. Tal como lo veían los maestros, ser un gran profesor quiere decir «algo». No estaban seguros de que ser un gran administrador implicara un peso equivalente.

CONCLUSIÓN

Las perspectivas que los profesores de Vista City trasladaban a su trabajo revelan la insuficiencia de las opiniones actuales sobre la carrera. Para incluir realmente a las mujeres, el concepto de carrera debe describir los patrones de las biografías tanto de las mujeres como de los hombres. Para esto, los datos deben ser generados tanto por las

experiencias de los hombres como de las mujeres, no de la vida de los hombres aplicada a las mujeres. Como sugiere el estudio de las biografías de estas maestras, no podemos tener un modelo de carrera que describa sólo a aquellas mujeres que combinan una gran ambición con la voluntad de desafiar las normas sociales.

Para revisar nuestra opinión, necesitamos una aproximación a las biografías laborales de mujeres más allá de los confines del concepto de la carrera masculina. Como sugieren las biografías presentadas, el papel de padre debe ser concebido como compatible, no como competitivo, con el papel laboral. Ser padre no se opone a poseer un compromiso laboral. Por decirlo de otro modo, cuando las mujeres eligen dar a luz y criar a sus hijos, eso no significa una ausencia de compromiso laboral. Si es necesario explicar por qué algunas mujeres vacilan en dejar a sus hijos, más bien deberíamos considerar las consecuencias estructurales de la desatención social a las necesidades infantiles y el estigma que se da al cuidado diario.

Un concepto revisado de carrera debe reflejar un concepto alternativo de éxito, aquél que no equipare éxito en la vida con éxito en el trabajo. Las críticas de las perspectivas actuales de la sociología del trabajo han advertido acertadamente que utilizar un modelo de género para las mujeres (en el cual se estudian los compromisos laborales y familiares) y un modelo laboral para los hombres (donde sólo son relevantes los temas laborales) conduce a la desigualdad social⁴⁰. Si las carreras son esencialmente masculinas, entonces las mujeres que tienen hijos pueden no ser tomadas nunca en serio. Debemos defender modelos diversos en vez de un concepto de trayectorias profesionales estrechamente definido. Que pensemos en el nacimiento y crianza de los niños como un desvío, revela las limitaciones de aplicación del concepto a la sociedad contemporánea. Se considera que, incluso las mujeres con profesiones con un estatus más elevado que la enseñanza, que hacen un alto en sus carreras para tener hijos, se toman su trabajo con «menos seriedad» que aquéllas que no interrumpen sus carreras. El término «interrupción de carrera» sugiere un desvío del modelo ideal creado alrededor de «mecanismo de relojería de las carreras de los hombres»⁴¹. Desde este punto de vista, las mujeres ocupan puestos de trabajo, pero no desarrollan una trayectoria profesional. Un modelo de carrera genérico debe, desde la base, tener en cuenta igualmente las vidas tanto de los hombres como de las mujeres.

REFERENCIAS

¹Una versión más actual fue presentada en el 10^o *Research on Women and Education Conference* CSU/Long Beach, California, noviembre 1984, y en la reunión anual de AERA, Chicago 1985. La investigación aquí presentada fue dirigida con ayuda del *Nacional Institute of Education*. Las opiniones expresadas, sin embargo, son mías. Me gustaría agradecer a Rally Gregory Kohlsdet, Douglas Biklen, Carol Shakeshaft y a los anónimos revisores de *Issues in Education* por sus comentarios de ayuda en las primeras versiones.

²Ver, por ejemplo, The National Commission on Excellence in Education, *A Nation At Risk: The Imperative for Educational Reform* (Washington: U.S. Department of Education, 1983); Ernest Boyer, *High School: A Report on Secondary Education in America* (Princeton, NJ: Carnegie Foundation for the Advancement of Teaching, 1983); y John Goodlad, *A Place Called School: Prospects for the Future* (New York: McGraw-Hill, 1983).

³Mary Kay Tetreault y Patricia Schmuck, «Equity As An Elective: An Analysis of Selected Educational Reform Reports and Issues of Gender», *Issues in Education* 3 (1985): 45–67.

- ⁴Philip Schlecty y V. Vance, «Recruitment, Selection and Retention: The Shape of the Teaching Force» «Research on Teaching: Implications for Practice», Warrentown, VA, February 1982).
- ⁵Burton Bledstein, *The Culture of Professionalism* (New York: Norton, 1976), 172. Harold Walensky, «The Professionalization of Everyone», *American Journal of Sociology* 70 (1964): 137–158.
- ⁶Un empleo alternativo de este concepto ha sido el empleo de carrera para describir un cambio en la vida de las personas. Con este uso se refiere a diversas posiciones, etapas y formas de pensar que la gente atraviesa en el transcurso de sus vidas. Destaca las perspectivas de vida de los participantes.
- ⁷Oswald Hall, «Stages of A Medical Career», *American Journal of Sociology* 26(1948): 524–538; Dan Lortie, «Laymen to Lawmen: Law Schools, Careers and Professional Socialization», *Harvard Educational Review* 29 (1959): 352–369; y L. Wilson, *The Academic Man* (New York: Oxford University Press, 1942).
- ⁸Dan Lortie, *School-teacher* (Chicago: University of Chicago, 1975).
- ⁹M. Blaxall y B. Reagan, eds., *Women and the Workplace: The Implications of Occupational Segregation*, *SIGNS* 1, pt. 2 (1976); y M. S. Larson, *The Rise of Professionalism* (Berkeley: University of California, 1977).
- ¹⁰Amitai Etzioni, ed., *The Semi-Professions and Their Organization* (New York: Free Press, 1969); y Blanche Geer, «Teaching», *International Encyclopedia of the Social Sciences* 15 (1968): 560–565.
- ¹¹Ver Veronica Nieva y Barbara Gulick, *Women and Work* (New York: Praeger, 1981).
- ¹²W. S. Mason, *The Beginning Teacher: Status and Career Orientations* (Washington DC: Government Printing Office, 1961).
- ¹³B. G. Bourne y N. J. Wilker, «Commitment and the Cultural Mandate: Women in Medicine», in *Women and Work*, eds. Rachel Kahn-Hut, Arlene Kaplan Daniels and Richard Colvard (New York: Oxford University Press, 1982): 111–122.
- ¹⁴Mary Frank Fox y Sharlene Hesse-Biber, *Women at Work* (Palo Alto, CA: Mayfield, 1984).
- ¹⁵Kahn-Hut, Daniels y Colvard, *Women and Work*, 2.
- ¹⁶Rose Coser y Gerald Rokoff, «Women in the Occupational World: Social Disruption and Conflict», *Social Problems* 18 (1970): 534–554.
- ¹⁷Es importante recordar que no se trata de categorías rígidas. La progresión no es semejante para todas las profesiones.
- ¹⁸Lortie, *School-teacher*.
- ¹⁹Howard S. Becker, «The Career of the Chicago Public School Teacher», *American Journal of Sociology* 57 (1952).
- ²⁰Lortie, *School-teacher*, 89.
- ²¹Ver W. S. Mason, R. J. Dressel, y R. K. Bain, «Sex Role and the Career Orientations of Beginning Teachers», *Harvard Educational Reviews* 29 (1959): 370–383; Robert Dreeben, *The Nature of Teaching* (Glenview, IL: Scott, Foresman, 1970); Blanche Geer, «Occupational Commitment and The Teaching Profession», *The School Review* 74 (1966); Oswald Hall, «The Social Structure of the Teaching Profession», en *Struggle for Power in Education*, eds. F. W. Lutz and J. J. Azzarelli (New York: Center for Applied Research in Education, 1966), 35–48; Dan Lortie, *School-teacher*; J. D. Grambs, «The Roles of the Teacher», en *The Teacher's Role in American Society*, ed. T. Stiles, 14th Yearbook of the John Dewey Society, New York; R. L. Simpson y I. H. Simpson, «Women and Bureaucracy in the Semi-Professions», en *The Semi-Professions and Their Organization*, 196–255; y M. G. Sobel, «Commitment to Work», en *Working Mothers*, eds. L. W. Hoffman y F. I. Nye (San Francisco: Jossey-Bass, 1975): 63–80.
- ²²Sari Knopp Biklen y Charol Shakeshaft, «The New Scholarship on Women», en *Handbook for Achieving Sex Equity in Schools*, ed. Susan Klein (Baltimore: Johns Hopkins Press, 1985), y Catharine Stimpson, «The New Scholarship about Women: The State of the Art», *Annals of Scholarship* 2 (1980): 2–14.
- ²³Ver Kahn-Hut, Daniels y Colvard, *Women and Work*; Cynthia F. Epstein, «Sex Role Stereotyping, Occupations and Social Exchange», *Women's Studies* 3 (1976): 185–194; Rosabeth M. Kanter, «The Impact of Hierarchical Structures on the Work Behavior of Women and Men», *Social Problems* 23 (1976): 415–430; Judith Long Laws, «Work Aspirations of Women: False Leads and New Starts», *Signs* 1:3, pt. 2 (1976): 33–49; y Coser y Rokoff, «Women in the Occupational World».
- ²⁴Ver Joan Acker, «Issues in the Sociological Study of Women's Work», en *Women Working*, eds. A. H. Stromberg y S. Harkness (Palo Alto, CA: Mayfield, 1978): 134–161. Cada vez aparecen más estudios que examinan cómo mujeres que enseñan construyen su propia experiencia. Ver, por ejemplo: Dee Spencer-Hall, *Teachers as Persons: Case Studies of the Lives of Women Teachers* N.I.E., Julio 1982; Margaret Nelson, «From the One-Room Schoolhouse to the Graded School: Teaching in Vermont, 1910–1950», *Frontiers* 7 (1983): 14–20; Richard Quantz, «Teachers as Women: An Ethnohistory of the 1930s», Reunión anual de American Educational Research Association, New York, Marzo 1982; Polly

Kaufman, *Women Teachers on the Frontier* (New Haven, CT: Yale University Press, 1984); Boston Women Teachers Group, «A Study of the Effect of Teaching on Teachers», Reunión anual de American Educational Research Association, Boston, 1980; y Michael Apple, «Work, Gender and Teaching», *Teachers College Record* 84 (1983): 611–628.

²⁵Judith Agassi, «The Quality of Women's Working Life», *The Quality of Working Life*, eds. L. Davis y A. Cherns, vol. 1 (New York: New York Free Press, 1975), 280–298; Sobel, «Commitment to Work»; Fox y Hess-Biber, «Women at Work», y Nieva y Gulick, «Women at Work».

²⁶Kanter, «The Impact of Hierarchical Structures».

²⁷Ver, por ejemplo, Constantina Safilios-Rothschild, «Towards the Conceptualization and Measurement of Work Commitment», *Human Relations* 24 (1971): 489–493, and Sobel, «Commitment to Work».

²⁸R. L. Feldberg y E. R. Glenn, «Male and Female: Job Versus Gender Models in the Sociology of Work», *Social Problems* (1979): 524–538.

²⁹Nona Glazer, «Overworking the Working Woman: The Double Day in a Mass Magazine», *Women's Studies International Quarterly* 3 (1980): 79–93.

³⁰Sari Knopp Biklen, *Teaching as an Occupation for Women: A Case Study on an Elementary School*. N.I.E. under Grant no. NIE-G-81-007.

³¹Robert Bogdan y Sari Knopp Biklen, *Qualitative Research for Education* (Boston: Allyn y Bacon, 1982). Ver también, Judith P. Goetz y Margaret Lecompte, *Ethnography and Qualitative Design in Educational Research* (Orlando, FL: Academic Press, 1984).

³²Ibid.

³³Los nombres de las escuelas y de los profesores han sido modificados para proteger su anonimato.

³⁴Para aceptar esta posición, ella tuvo que rechazar acompañar a su marido a Europa durante su año sabático. Esto causó ciertas disputas conyugales temporales.

³⁵El patrón de Kate se reflejó en las vidas de otros profesores casados en *Vista City Elementary*. Jessica Bonwit, por ejemplo, abandonó la enseñanza a tiempo completo durante seis años mientras sus tres hijos eran muy pequeños. Durante este periodo, sin embargo, no paró de trabajar. Comenzó un programa de clases particulares en su casa para niños de seis a ocho años. Christine Bart describió su historial de trabajo como «continuo» aunque no en trabajos permanentes.

³⁶Sugiero que este ejemplo es el típico para solucionar los problemas de muchos profesores. A menudo, buscaban la resolución individual antes que un cambio de la situación.

³⁷A lo mejor, Jessica estaba preparada, en cierto modo, para un matrimonio de este tipo. Su propia educación no destacó su valía en un mundo de hombres: *Yo crecí en una familia italiana en la que los chicos eran mucho más valorados que las chicas. Así que a mi hermano siempre le escuchaban con mucha atención. Pero a mí me escuchaban a medias. Era como: «¿qué estabas diciendo?».* Así que me acostumbré a hablar las cosas conmigo misma.

³⁸Incluso la directora explicó su desgana a considerar su solicitud como superintendente. La alejaría demasiado de los niños.

³⁹Al mismo tiempo, los profesores que estaban desocupados estaban temerosos por la inseguridad de su situación. Como desocupada, Jessica Bonwit dijo el último día de clase cuando no estaba segura de si sería posible volver el año siguiente: «Las circunstancias de la situación laboral me resultan muy desmoralizadoras. Creo que para los profesores resulta demoledor no saber dónde van a estar el año siguiente. Es un modo terrible de tratarnos». La estructura ocupacional era lamentable.

⁴⁰Feldberg and Glenn, «Male and Female».

⁴¹Arlie Hochschild, «Inside the Clockwork of Male Careers», *Women and the Power to Change*, ed. Florence Howe (New York: McGraw-Hill, 1971): 47–80.

Source: «Can Elementary Schoolteaching Be a Career? A Search for New Ways of Understanding Women's Work», Sari Knopp Biklen, *Issues in Education*, Vol. III, No. 3, Winter 1985. Copyright © 1985 por la American Educational Research Association.

Conclusiones. Las conclusiones incluyen, normalmente, un nuevo planteamiento del enfoque inicial del estudio, y cómo los datos resultantes y los análisis inciden en ese enfoque. Se pueden elaborar las implicaciones que los resultados conlleven, así como implicaciones para la investigación posterior.

Guías para la evaluación investigación cualitativa

Para entender la investigación cualitativa es necesario leer cuidadosamente el informe completo. Así es como será capaz de identificarse con los investigadores y entender cómo han llegado a sus conclusiones. El proceso por el cual ocurre esto es importante y para comprenderlo es necesario leer desde el principio hasta el final. Como con los estudios cuantitativos, se deben realizar ciertas preguntas sobre el informe para juzgar su calidad.

Introducción

1. ¿Se ha indicado claramente el enfoque, el propósito o el tema del estudio?
2. ¿Hay situaciones o problemas que dirigen al foco del estudio? ¿Hay base lógica para el estudio? ¿Es evidente que el estudio es importante?
3. ¿Existen antecedentes en la investigación y una teoría que ayuden a perfeccionar las preguntas de investigación?
4. ¿La introducción contiene una visión general del diseño?
5. ¿La revisión de la bibliografía es pertinente con el tema de la investigación? ¿Está la bibliografía descrita y analizada?

Metodología

1. ¿Los escenarios están descritos para identificar lo que tienen de único o de característico?
2. ¿Cómo se estableció la entrada inicial en el campo?
3. ¿Cómo se explicó la presencia del investigador en el campo a los demás? ¿Cuál era el papel del investigador?
4. ¿Quién fue observado? ¿Durante cuánto tiempo? ¿Cuánto tiempo llevó la recogida de datos?
5. ¿Informa el investigador de restricciones para acceder a los datos?
6. ¿Son representativos los datos de la conducta en circunstancias normales?
7. ¿Se reconocen las limitaciones del diseño?

Resultados e interpretaciones

1. ¿Se presentan con claridad las perspectivas de los diferentes participantes? ¿Están entrecomillados los comentarios de los participantes?
2. ¿Se proporciona la información contextual de las afirmaciones de los participantes?
3. ¿Se presentan diversas perspectivas?
4. ¿Están bien documentados los resultados? ¿Se ilustran por medio de los resultados las afirmaciones e interpretaciones?
5. ¿Queda claro que los investigadores creen en los datos indicados? ¿Las opiniones personales se mantienen separadas de los datos?

6. ¿Son razonables las interpretaciones? ¿Se reconocen las ideas previas y los sesgos del investigador?

Conclusiones

1. ¿Tienen las conclusiones consistencia lógica con los resultados?
2. ¿Se indican las limitaciones del diseño y el enfoque de investigación?
3. ¿Se han indicado las implicaciones de los resultados?

RESUMEN

Este capítulo ha proporcionado una visión general de la terminología común de los tipos de investigación con referencia a la modalidad de investigación, las técnicas utilizadas para la recogida de datos y el formato estándar de los artículos publicados. Los puntos principales de este capítulo son los siguientes:

1. La modalidad de investigación es la aproximación general y la preferencia metodológica empleada en el estudio.
2. El diseño de investigación es el plan general del estudio, incluyendo cuándo, de quién y cómo se recogen los datos.
3. En la investigación experimental, el investigador estudia las relaciones de causa y efecto manejando un factor y observando cómo este factor se relaciona con el resultado del estudio.
4. La investigación experimental se caracteriza por la asignación aleatoria de sujetos a grupos, el control sobre los factores que pueden desvirtuar el estudio y decisiones como qué tratamiento recibe cada grupo.
5. La investigación semiexperimental investiga la causalidad sin asignación aleatoria y control completo.
6. La investigación de caso único investiga la relación causal entre un factor y la conducta de un único individuo.
7. «No experimental» es un término genérico referido a una investigación en la que no existe control directo sobre la causalidad. Las modalidades de investigación no experimentales pueden clasificarse como descriptivas, comparativas, correlacionales, mediante encuesta o *ex post facto*.
8. Las técnicas para recoger la información incluyen aproximaciones cuantitativas y cualitativas. Las técnicas cuantitativas, como cuestionarios y tests, utilizan números como datos, mientras que las técnicas cualitativas, como la etnografía, emplean descripciones narrativas.
9. Las modalidades cualitativas interactivas utilizan la recogida de datos cara a cara para elaborar la comprensión en profundidad de las perspectivas de los participantes.
10. Una investigación etnográfica consiste en una descripción detallada e interpretación de una cultura o sistema.

11. Un estudio fenomenológico describe los significados de una experiencia vivida desde las perspectivas de los informantes.
12. Un estudio de caso investiga, utilizando múltiples fuentes de datos, un sistema definido a través del tiempo.
13. Para desarrollar conceptos detallados o propuestas sobre fenómenos particulares se recurre a la teoría fundamentada.
14. Los estudios críticos subrayan la subjetividad del conocimiento y las perspectivas actuales de la teoría crítica, feminista y postmoderna.
15. Las modalidades cualitativas no interactivas o investigación analítica investigan conceptos y sucesos a través del análisis de documentos.
16. El formato de los estudios cuantitativos sigue una secuencia establecida de antemano con secciones similares. En los estudios cualitativos, el formato variará pero, normalmente, incluirá una introducción y revisión de bibliografía, metodología, resultados e interpretaciones y conclusiones.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La investigación experimental contiene todas las características siguientes *excepto*...
 - a. un grupo de control o de comparación.
 - b. asignación aleatoria de sujetos.
 - c. entrevistas intensivas a los sujetos.
 - d. manipulación de sujetos.
2. Los diseños semiexperimentales difieren de los diseños experimentales en...
 - a. el grado de control.
 - b. el empleo de grupo de comparación o control.
 - c. la asignación aleatoria.
 - d. la a y la c son correctas.
3. En la investigación no experimental el investigador puede hacer todo lo siguiente *excepto*...
 - a. hacer inferencias de una muestra a una población.
 - b. describir las condiciones existentes.
 - c. establecer relaciones causa-efecto.
 - d. predecir un fenómeno a partir de otro.
4. Un estudio etnográfico se puede clasificar como...
 - a. fenomenológico.
 - b. no interactivo.

- c. interactivo.
 - d. teoría fundamentada.
5. En una investigación correlacional los investigadores estudian...
- a. las condiciones existentes de una manera descriptiva.
 - b. relaciones causa-efecto.
 - c. el mismo suceso en dos momentos en el tiempo.
 - d. el grado de relación entre dos fenómenos.
6. Las observaciones estructuradas están en la misma categoría que...
- a. la investigación aplicada.
 - b. la investigación descriptiva.
 - c. la investigación evaluativa.
 - d. ninguna de las anteriores.
7. Las técnicas cualitativas de recogida de datos se distinguen de las técnicas cuantitativas por...
- a. el diseño de investigación utilizado.
 - b. la función de la investigación.
 - c. el empleo de palabras en lugar de números
 - d. el suministro de resultados más significativos.
8. La investigación fenomenológica a menudo se caracteriza por...
- a. entrevistas intensivas.
 - b. notas de campo abundantes.
 - c. hipótesis claramente definidas.
 - d. observaciones no intrusivas.
9. El resumen de un artículo contiene...
- a. una descripción completa de los sujetos.
 - b. recomendaciones para estudios posteriores.
 - c. el propósito del estudio.
 - d. la a y la c son correctas.
10. La sección de metodología de un artículo de investigación cuantitativa contiene...
- a. sujetos, materiales, pruebas.
 - b. participantes, pruebas, procedimiento.
 - c. sujetos, pruebas, análisis de datos.
 - d. participantes, sujetos, procedimientos.
11. En la investigación cuantitativa la hipótesis sigue a...
- a. el problema de investigación.
 - b. la revisión de bibliografía.
 - c. la introducción.
 - d. el resumen.
12. El propósito de la sección de metodología en los estudios cualitativos es...

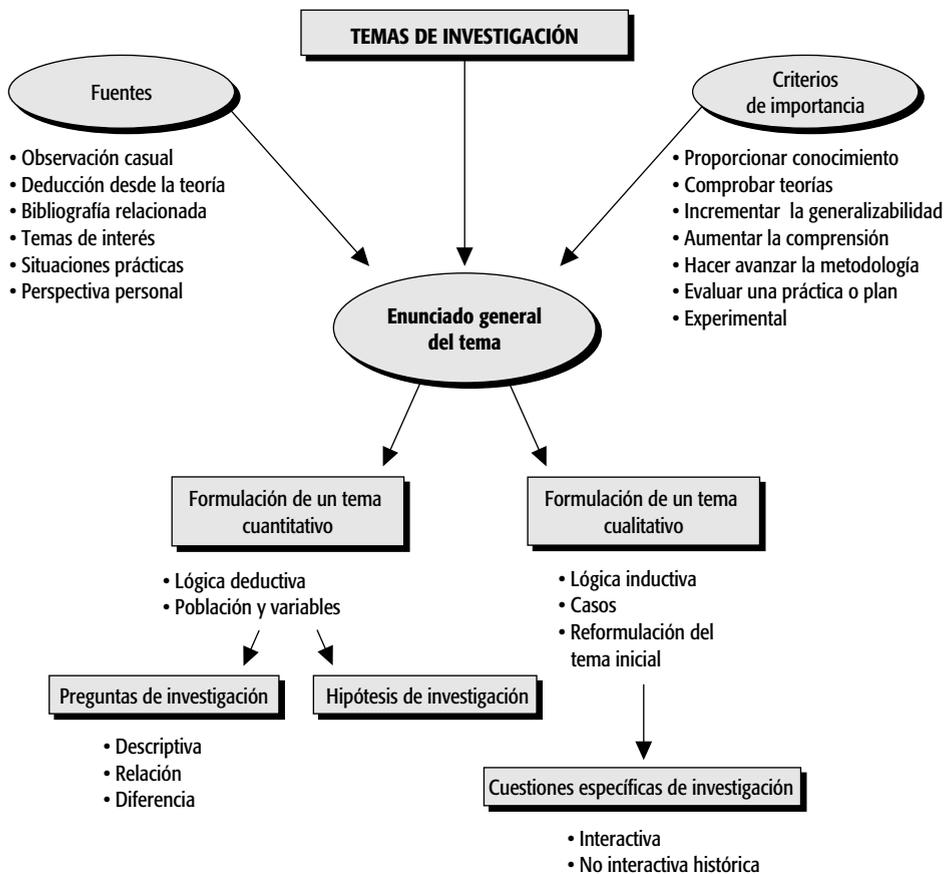
- a. proporcionar un contexto general para el estudio.
- b. resumir el diseño del estudio.
- c. mostrar cómo el investigador hará las interpretaciones.
- d. presentar el enfoque del estudio.

Problemas aplicados

1. Clasifique cada estudio descrito a continuación con respecto a su diseño de investigación: experimental, no experimental, cuantitativo o cualitativo. Puede ser válido más de un tipo.
 - a. una investigación piloto de la validez del test de la brazada en el estilo de espalda para identificar problemas en los nadadores.
 - b. una comparación del efecto de dos programas de lectura en clases de cuarto curso en Kalamazoo.
 - c. una investigación de la estructura de las actitudes de alumnos universitarios.
 - d. el efecto de recompensas extrínsecas en la motivación de niños asignados de forma aleatoria a grupos de juego.
 - e. una encuesta sobre las actitudes de los directores hacia la negociación colectiva.
 - f. un estudio de la efectividad relativa de diferentes técnicas de orientación que han utilizado los orientadores durante los últimos cinco años.
 - g. una investigación de la diferencia de asistencia entre dos institutos con diferentes estilos de dirección.
 - h. un estudio posttest del efecto de frases humorísticas en la comprensión de dos grupos de niños.
 - i. un estudio del significado de la paga extra por méritos para los profesores.
2. Lea un artículo de investigación cuantitativa e identifique las frases que correspondan a las secciones estándar citadas a continuación.
 - a. resumen
 - b. introducción
 - c. enunciado del problema de investigación
 - d. revisión de la bibliografía
 - e. enunciado de las hipótesis de investigación
 - f. sujetos
 - g. instrumentos/ pruebas
 - h. procedimientos
 - i. resultados
 - j. discusión, implicaciones, conclusiones

3. Lea un artículo de investigación cualitativa e identifique las secciones principales:
 - a. introducción
 - b. metodología
 - c. resultados e interpretaciones
 - d. conclusiones
4. Describir la investigación en los problemas 2 y 3 con respecto a la modalidad de investigación y a la técnica de recogida de datos.

Temas de investigación: enunciados, preguntas e hipótesis



PALABRAS CLAVE

tema de investigación

cuantitativa:

constructo

variable

variable categórica

variable continua o medida

variable dependiente

variable independiente

variable experimental

o manipulada

variable predictora

variable criterio

definición operacional

hipótesis de investigación

cualitativa:

caso

registros cualitativos de campo

descripciones narrativas

resúmenes sintetizados

problemas anunciados

significación del problema

Este capítulo presenta el que, probablemente, sea el aspecto más difícil de la investigación: la formulación de un tema de investigación claro, conciso y manejable. El enunciado del tema de investigación resulta crucial porque informa del enfoque y de la importancia del problema, el contexto educativo y su alcance y el marco para dar a conocer los resultados. Formular un tema de investigación requiere un trabajo preliminar considerable y la selección de la lógica apropiada, ya sea inductiva o deductiva.

Este capítulo, también, aborda las fuentes más frecuentes que se emplean para generar temas de investigación y los criterios para juzgar su importancia o su significado. Las formulaciones cuantitativas del tema emplean el razonamiento deductivo para seleccionar los constructos, las variables y los instrumentos. Los problemas cualitativos requieren razonamiento inductivo. Los temas cualitativos se reformulan mientras el investigador los elabora, a partir de las observaciones de situaciones sociales de sucesos actuales o pasados de un caso seleccionado, y después los relaciona con fenómenos más amplios. También haremos referencia a los criterios para juzgar la adecuación del enunciado del tema.

LA NATURALEZA DE LOS TEMAS DE INVESTIGACIÓN

Los investigadores pueden hacerse muchas preguntas sobre las teorías y prácticas educativas. Considere la siguiente serie de preguntas:

- ¿Cuál es el estado actual de las opiniones o funcionamiento de un grupo?
- ¿Cuáles son los efectos de una práctica, innovación o política específica?
- ¿Cómo influyen los sucesos históricos, las leyes o las políticas en los asuntos educativos actuales?
- ¿Cómo enseña biología la Sra. Kay a los alumnos con bajo rendimiento?
- ¿Cuáles son las posiciones teóricas que guían las prácticas diarias y la planificación a largo plazo?
- ¿Por qué Ellen permanece callada en una tarea de lectura en pequeño grupo y muestra una conducta revoltosa cuando los alumnos eligen las actividades y los centros de aprendizaje?
- ¿Cuál es la mejor forma de orientar el trabajo propio?

Preguntas como éstas constituyen el paso inicial en una investigación.

Sin embargo, algunas preguntas que son interesantes para un grupo o un individuo pueden no suponer temas de investigación. El sustantivo *tema* tiene significados convencionales y técnicos. En el sentido convencional, un tema es un conjunto de condiciones que necesitan discusión, una decisión, una solución o información. Un **tema o problema de investigación** implica la posibilidad de indagación empírica, esto es, una recogida y análisis de los datos.

Explicaciones sobre cómo hacer algo, propuestas ambiguas y preguntas de valoración no son temas de investigación *per se*. Cuestiones como: «¿cómo podemos conseguir la igualdad de oportunidades?» o «¿cómo podemos prevenir el fracaso escolar?» son preguntas acerca del *cómo*. Formulaciones como: «las instituciones democráticas son una manifestación natural de la cultura norteamericana» resultan demasiado ambiguas o amplias para ser objeto de una investigación. Las preguntas valorativas cuestionan cuál de dos o más elementos es bueno o malo, deseable o no deseable, mejor o peor, debería o no debería hacerse. Las preguntas de valoración, como las que hemos indicado, no pueden ser investigadas experimentalmente. Y aunque las preguntas acerca del «cómo», las formulaciones ambiguas y las preguntas de valoración son importantes para los directores de colegios, padres, profesores, filósofos y dirigentes políticos, están enunciadas de ese modo, más allá de la investigación. Sin embargo, mediante el planteamiento de estas preguntas puede surgir un tema objeto de investigación.

En contraste con un problema práctico, un tema de investigación se formula para indicar la necesidad de una investigación empírica. Los temas o problemas de investigación cuantitativos pueden expresarse como preguntas o hipótesis. Veamos algunos ejemplos. «¿Cuáles son las actitudes de los padres hacia la política de repetición de curso de una escuela?». «¿Existen diferencias en los periodos de desarrollo adulto de Levinson entre grupos de alumnos y alumnas licenciados?». «Existe una relación positiva entre la asistencia a preescolar y la madurez social en alumnos de primaria, es decir, A se relaciona con B». «¿Influyen la aptitud académica, la autoestima y el nivel de aspiración en el

rendimiento académico, es decir, A, B y C se relacionan con D?». Cada uno de estos enunciados implica recogida y análisis de datos.

Los temas de investigación cualitativos se expresan como enunciados de investigación o preguntas, pero *nunca* como hipótesis. Una hipótesis de investigación implica razonamiento deductivo; la investigación cualitativa utiliza principalmente razonamiento inductivo para sugerir una interpretación de una situación particular o de un periodo histórico. Los problemas cualitativos, normalmente, se expresan de forma más amplia que los problemas cuantitativos empleando términos como *qué, cómo y por qué*. Los problemas cualitativos indican la situación o el contexto de tal manera que delimitan el problema. Un tema cualitativo puede ser el estudio de una situación específica, una persona, un estado o un periodo histórico. Algunos ejemplos de preguntas de investigación cualitativas son: «¿cómo ayuda la Sra. Jackson, una profesora de primaria, a Jim, un estudiante de magisterio durante sus prácticas?» o «¿qué significa ser un ‘padre soltero’ para el Sr. Strong con dos niños por criar: cómo afecta ser padre soltero a su papel de padre y a otros aspectos de su vida?».

Los estudios cualitativos también examinan el pasado a través de documentos históricos o jurídicos. Estas preguntas de investigación, normalmente, se redactan en tiempo pasado. Por ejemplo, «¿cuándo y por qué se requirió que los profesores se titularan en el estado de Vermont?». «¿Cómo han cambiado los requisitos para la titulación de maestros desde entonces?». «¿Por qué se realizaron esos cambios?». Los temas de investigación cualitativa pueden redactarse como enunciados de investigación tal como: «el propósito de este estudio es examinar y analizar los fundamentos jurídicos defendidos por las cortes federales y estatales en los casos relacionados con los trastornos de los alumnos desde 1952 hasta 1999 para proporcionar una definición judicial del término ‘trastornos de los alumnos’».

Fuentes de los temas

Los temas, inicialmente, se identifican como temas generales. Después de mucho trabajo preliminar, el tema general se enfoca como un problema de investigación específico. ¿Dónde se inicia el descubrimiento de los temas generales? Las fuentes más comunes son las observaciones casuales, las deducciones a partir de la teoría, la revisión de la bibliografía, los temas sociales de actualidad, las situaciones prácticas y las experiencias y perspectivas personales. Examinaremos los tipos de estudios que sugieren estas fuentes.

Observaciones casuales. Las observaciones casuales son fuentes ricas de preguntas y corazonadas. Las decisiones, normalmente, se basan en los posibles efectos de las prácticas educativas en los alumnos, la plantilla o en la comunidad sin poseer datos empíricos. Las preguntas de investigación pueden ser sugeridas

por la observación de ciertas relaciones para las que no existe una explicación satisfactoria, las formulas rutinarias de hacer las cosas que están basadas en la autoridad o la costumbre, que carecen de evidencia empírica, o en cambios tecnológicos e innovaciones que necesitan una confirmación a largo plazo. Tales estudios pueden resolver un problema práctico, proponer una nueva teoría o identificar variables que todavía no aparecen en la bibliografía.

Deducciones desde la teoría. Las deducciones a partir de la teoría pueden sugerir temas de investigación. Las teorías son principios generales cuya posible aplicación a problemas educativos específicos es desconocida hasta que no es puesta a prueba de forma experimental. La validez y el alcance de una variedad de teorías podrían ser puestas a prueba bajo condiciones educativas. Tales estudios podrían verificar la utilidad de una teoría para explicar las situaciones educativas.

Bibliografía relacionada. La bibliografía relacionada puede sugerir la necesidad de refutar un estudio con o sin variación. La repetición de un estudio puede incrementar la generalizabilidad y la validez de los resultados anteriores. En muchos casos, no es posible la asignación aleatoria de sujetos, lo que limita la *generalizabilidad* de los resultados. Pero, en tanto los experimentos se repitan en diferentes momentos y en diferentes entornos con resultados similares, los investigadores pueden confiar más en los hallazgos obtenidos. Citar la bibliografía relacionada permite a los estudios cualitativos *ampliar la comprensión empírica* a otras situaciones.

Temas sociales y políticos de actualidad. Los temas sociales y políticos de actualidad de la sociedad norteamericana son frecuentes en la investigación educativa. El movimiento de las mujeres planteaba cuestiones sobre la igualdad de sexos en general y sobre los estereotipos sexuales en los materiales y prácticas educativas. Los derechos civiles llevaron a investigar la educación de los niños pertenecientes a minorías y los efectos de la segregación en las actitudes raciales, las relaciones entre razas, la autoestima, el rendimiento académico y asuntos similares. Las políticas de inmigración recientes sugieren preguntas de investigación sobre el multiculturalismo en la educación

Situaciones prácticas. Debido a que los que toman las decisiones en un lugar determinado necesitan información, las situaciones prácticas pueden sugerir estudios de planes y de evaluación. Aunque un tema de investigación no se define como una cuestión de valores, la información que genera se emplea para la toma de decisiones basada en valores. Las preguntas para tal investigación pueden estar centradas en las necesidades educativas; en la información para la planificación, desarrollo e implantación de programas o la eficacia de una práctica.

Experiencia personal e *insight*. Las experiencias personales y el *insight* (toma de conciencia o conocimiento súbito) pueden sugerir temas de investigación que deberían examinarse más profundamente a través de metodologías cualitativas.

Por ejemplo, un profesor que ha trabajado con niños superdotados puede reconocer más fácilmente los significados de una situación con este tipo de niños que un etnógrafo que no ha tenido contacto con ellos. Un historiador que ha trabajado como director de un departamento universitario podría establecer mayor empatía con un personaje histórico que hubiera sido rector de una universidad. La habilidad para establecer empatía y reconocer los significados sutiles en una situación es importante en la mayoría de las investigaciones cualitativas.

Enunciados formales del tema

Los investigadores utilizan enunciados formales del tema para orientar su investigación. Los enunciados presentan al lector la importancia del problema, lo sitúan en un contexto educativo y le proporcionan el marco para presentar los resultados. El planteamiento del tema orienta al lector sobre la trascendencia del estudio y las preguntas de investigación o hipótesis que lo siguen.

Enfoque, contexto educativo e importancia del tema. Ante un enunciado de tema escrito correctamente, el lector no queda intrigado sino que comprende inmediata y directamente el enfoque general, el contexto educativo y la trascendencia del problema. Por ejemplo, en los primeros párrafos de «¿Quién dirigirá? Los 10 principales factores que influyen en los profesores para convertirse en directores» (Cooley y Shen, 1999), al lector se le informa sobre la escasez del número de puestos directivos en los colegios y la resistencia de los profesores para formar parte de la administración. Se llevó a cabo el estudio para identificar los factores que influían en la decisión de los profesores para solicitar un puesto de gestión administrativa (ver extracto 3.1).

EXTRACTO 3.1

ENUNCIADO DEL PROBLEMA: MARCO

¿Quién dirigirá los colegios durante el siglo XXI? Aproximadamente el 60% de los directores que actualmente ocupan puestos administrativos alcanzan la edad de jubilación el próximo año (2000) (Parkay y Currie, 1992). El problema está agravado por el descenso del número de profesores que buscan la cualificación administrativa y por el hecho de que muchos de los que están estudiando no se plantean ocupar un puesto administrativo después de terminar sus estudios (Jordon, 1994)... El rechazo del profesor a formar parte de la administración, añadido al número de jubilaciones de puestos en la dirección administrativa, supone un desafío importante para juntas, inspectores y comunidades...

Se realizó un estudio con 189 estudiantes de Master en Dirección Escolar de una universidad del medio oeste para ayudar a encontrar una solución a este problema. Los estudiantes completaron una encuesta identificando los factores que influían en su decisión de solicitar un puesto administrativo...

Los 10 factores principales que influían en los profesores para solicitar un puesto administrativo refieren una lista compleja de «deseos y necesidades» de esta nueva generación de directores de escuelas. Los factores enumerados en la tabla 1 presentan implicaciones para retener y reclutar directores de colegio.

Fuente: De Cooley, V. y Shen, J. (abril, 1999), «Who will lead. The top 10 factors that influence teachers moving into administration», *NAESP Bulletin*, 83 (606), 75-80.

En *The High School as a Social Service Agency: Historical Perspectives on Current Policy Issues* (El instituto como una agencia de servicio social: perspectivas históricas de temas políticos actuales) se replantea un enfoque, un contexto educativo y una importancia del problema diferentes. «El propósito de este ensayo es examinar históricamente la extensión de los servicios sociales a la juventud a través de la educación pública secundaria norteamericana... para arrojar luz sobre las difíciles elecciones que afrontan hoy en día los políticos» (Tyack, 1979, p. 45). En los dos primeros párrafos se le dice al lector que el contexto educativo se refiere a la educación pública secundaria norteamericana, que el concepto estudiado es la extensión de los servicios sociales, que el estudio es histórico, abarca desde 1890 hasta 1975 y que la comprensión de las razones de que tales programas fueran ampliados puede ayudar, en primer lugar, a los políticos a decidir qué funciones y servicios educativos recortar para cuadrar los presupuestos.

Un marco para los resultados y las conclusiones. El enunciado del tema proporciona, también, un marco para dar a conocer las conclusiones. A diferencia del diseño de investigación, en el que los investigadores definen qué métodos utilizarán para completar el estudio, el enunciado del tema simplemente indica lo que probablemente se necesita para realizar el estudio y explica que los resultados responderán a esta información. El extracto 3.1 ilustra el marco para los resultados de un estudio. Los diez factores que los profesores tenían en cuenta para solicitar puestos administrativos se presentaron y discutieron con las implicaciones para reclutar y mantener una nueva generación de directores escolares.

Los enunciados de problemas cualitativos también proporcionan el marco para dar a conocer los resultados e interpretaciones. Por ejemplo, en un estudio reciente, el problema consistía en describir «los resultados de la percepción de las madres solteras pobres con respecto a la ayuda de sus sistemas de apoyo para que puedan trabajar» (Wijnberg y Weinger, 1998). Los resultados se organizaron como primeras aspiraciones y su recuperación, apoyo y recursos percibidos de la red, estilos de hacer frente a la situación y apoyo social durante la enfermedad.

Los párrafos introductorios de un estudio son difíciles de escribir: se debe transmitir mucha información de forma resumida. Los investigadores reescriben,

varias veces, los párrafos cuando formulan la importancia del estudio y las preguntas de investigación. Incluso puede que escriban la versión final después de que el estudio esté totalmente terminado. Sin embargo, los investigadores, comienzan con un enunciado inicial del tema para guiar sus actividades.

FORMULACIÓN DEL PROBLEMA EN LA INVESTIGACIÓN CUANTITATIVA

El punto de partida para la definición de un tema de investigación es el planteamiento de preguntas sobre una cuestión relacionada con la educación, pero los temas de investigación educativa no son enunciados del problema en sí mismos. Un planteamiento de problema es más específico que un tema y delimita el alcance del tema de investigación.

Un investigador comienza con un tema general y lo reduce a un problema. Por ejemplo, el tema de la creación de leyes educativas podría centrarse en las políticas de la junta de dirección de la escuela. Éste todavía es un tema amplio y puede reducirse a ciertas políticas como la fiscal, estudiantil, de personal. Incluso, se puede enfocar más el tema: ¿está interesado el investigador en los antecedentes de las políticas, el proceso de elaboración de políticas o en sus consecuencias? ¿Existen repercusiones de las políticas y, en ese caso, sobre quién? Se pueden, pues, generar un buen número de preguntas a partir de un tema general: ¿qué influye en la formulación de políticas de la junta de dirección del colegio? ¿Cuál era la intención de la política de evaluación del profesorado y cómo se implementaron los procedimientos? ¿Qué efectos tiene la evaluación del profesorado en la moral de los profesores? ¿Qué efecto tiene la evaluación del profesorado en las actitudes y comportamientos de los administradores o directores? ¿Cuáles son las opiniones de los puestos directivos de las escuelas, padres y profesores hacia nuestra política de evaluación del profesorado? ¿Con qué frecuencia, en los últimos cinco años, la política de evaluación del profesorado en un colegio ha supuesto el cambio de comportamiento del profesor, la falta de renovación del contrato, la denegación del contrato fijo o procesos judiciales?

Suponga que el tema de interés es la enseñanza. De nuevo, un investigador podría preguntar cuestiones similares: ¿qué tipo de enseñanza? ¿Está centrada en los antecedentes, los procesos o las consecuencias? ¿Está enfocada en edades específicas o en alumnos con ciertas características? Un tema como el empleo de calculadoras en la enseñanza de matemáticas podría generar la siguiente pregunta: «¿existe diferencia entre alumnos comparables a causa del uso de calculadoras para el cálculo matemático?». Perciba que la pregunta falla al no especificar la frecuencia con la que se utilizan las calculadoras, cómo se determina el cálculo matemático o qué alumnos son comparables. Este problema sería más preciso

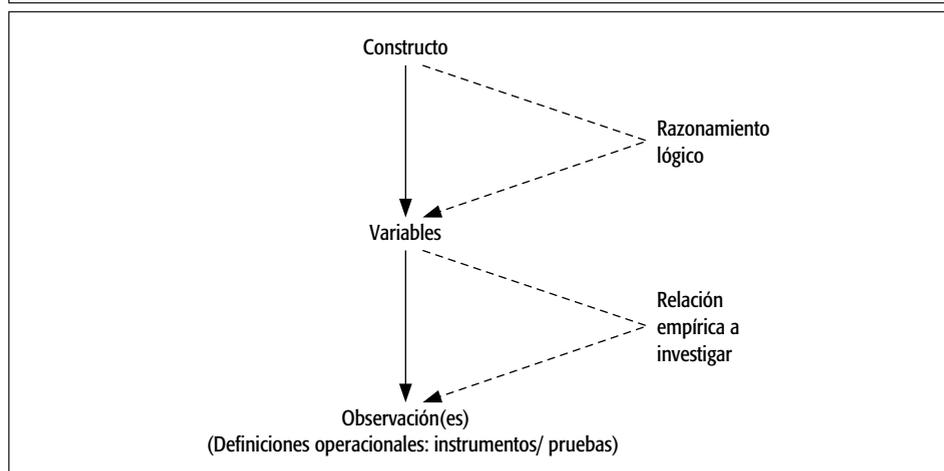
si se enunciara: «entre alumnos comparables de cuarto curso, ¿existen diferencias en las puntuaciones del test de rendimiento SRA de cálculo matemático entre los alumnos que han usado calculadoras en matemáticas durante un semestre y aquéllos que no lo han hecho?».

Para hacer el estudio manejable, los investigadores reducen sus temas a un problema particular. Si un problema es demasiado general, los resultados son difíciles de interpretar. Hay varias actividades que pueden ayudar a centrar un tema como problema. La lectura de la bibliografía auxiliar puede esclarecer el problema y reducirlo a posibles variables y preguntas. Hablar con aquéllos que podrían usar el estudio puede aclarar sus necesidades y preguntas. Una tormenta de ideas con otros que tienen experiencia en investigación o conocimiento especializado en el área del problema resulta beneficiosa. Por último, el investigador tiene que decidir sobre la selección de variables, la población y la *lógica para el problema*. Los enunciados iniciales del problema se revocan y se rescriben muchas veces para buscar el significado exacto de cada palabra y el razonamiento lógico a emplear durante la investigación. La lógica deductiva de los problemas cuantitativos se ilustra en la figura 3.1. Explicamos cada término y sus posibles significados en la siguiente sección.

La lógica deductiva de constructos, variables y definiciones operacionales

Para formular un problema, el investigador comienza con un constructo abstracto y, luego, determina si las variables seleccionadas se deducen lógicamente

FIGURA 3.1: LÓGICA DEDUCTIVA EN INVESTIGACIÓN CUANTITATIVA: CONSTRUCTOS, VARIABLES Y OBSERVACIONES



desde el constructo. El siguiente paso es seleccionar las observaciones razonadas para relacionarlas con las variables por deducción. Entonces las observaciones se definen operacionalmente a través de una prueba elegida por el investigador para medir el suceso. La figura 3.1 representa, de forma esquemática, la secuencia de los tres pasos del razonamiento deductivo desde los constructos abstractos hasta variables menos abstractas y, de ahí, a una serie de observaciones. Más adelante se explica el proceso con más detalle. Tenga en cuenta que la dirección de las flechas, para representar el razonamiento deductivo, es de arriba hacia abajo. En otras palabras, el investigador debe tomar varias decisiones para proponer un diseño predeterminado *antes* de la recogida de datos.

Constructos. En investigación, los conceptos abstractos de alto nivel se llaman **constructos**. El constructo expresa la idea tras el conjunto de casos particulares. A menudo, un constructo se deriva de una teoría. Por ejemplo, el constructo *ansiedad* se deriva de la teoría de la personalidad. Un **constructo** es una abstracción compleja que no es observable directamente. Ejemplos de constructos son: motivación, inteligencia, pensamiento, ansiedad, agresividad, autoestima, rendimiento y aptitud. Lo que es observable es el comportamiento o la respuesta que se presume es una consecuencia del constructo hipotetizado. Otra forma de definir un constructo es decir que podemos crear constructos combinando conceptos en patrones significativos. Conceptos como percepción visual, discriminación visual y auditiva, agudeza auditiva y lateralidad se combinan de forma significativa para sugerir el constructo *preparación para la lectura*. Normalmente se reconoce que el constructo *creatividad* consiste en flexibilidad, originalidad, elaboración y otros conceptos. Los constructos cambian su significado o son rechazados conforme se desarrollan las teorías o se acumula la evidencia empírica.

Debido a que los constructos son entidades intangibles y no son directamente observables, los investigadores emplean indicadores como medio de dar significado o clasificar la mayor parte de las particularidades del constructo. Los indicadores se denominan *variables*. Como mencionamos anteriormente, puede haber más de una variable para un constructo y más de un tipo de observación para una variable. El investigador, elige de forma deductiva, los indicadores más válidos.

Variables. Una **variable** es un suceso, categoría, comportamiento o atributo que expresa un constructo y, dependiendo de cómo se emplee en un estudio particular, posee valores diferentes. Hay varios tipos de variables. La **variable categórica** se emplea para diferenciar sujetos, objetos o entidades en dos o más categorías. La variable dicotómica es el tipo más simple de variable categórica y sólo tiene dos clases. Hombre-mujer, casado-soltero y aprobar-suspender son variables dicotómicas. Las variables categóricas pueden tener más de dos clases como por ejemplo nivel de ingresos, nivel educativo, nacionalidad o religión. Una **variable continua** es aquella en la que la propiedad o atributo de un objeto,

sujeto o entidad se mide numéricamente y puede tomar un número infinito de valores dentro de un rango. Peso, altura y edad son ejemplos comunes de variables continuas. En educación, rendimiento, habilidad verbal, aptitud y habilidades específicas son variables continuas habituales.

La investigación cuantitativa normalmente implica relaciones de variables categóricas y continuas. Algunas variables son antecedentes de otras variables. Los investigadores pueden saberlo por estudios previos o pueden hipotetizar de una teoría que una variable es un antecedente de otra. Por ejemplo, se puede decir que la inteligencia precede al rendimiento académico. Se asume que para rendir en la escuela, un niño necesita cierto grado de inteligencia. La variable *inteligencia* es un antecedente de la variable *rendimiento*. En cierto modo, el rendimiento es una consecuencia de la inteligencia; es decir, es dependiente del grado de inteligencia de un individuo. La variable que es consecuencia de algún fenómeno es el foco del estudio y, normalmente, es una variable continua. La variable antecedente puede ser tanto continua como categórica. Cada variable debería ser un fenómeno separado y distinto. La investigación educativa explora muchos factores como variables: variables del aula tales como estilos de enseñanza, patrones de interacción, nivel cognitivo de las preguntas; variables ambientales como nivel educativo de los padres, clase social, estructura familiar; y variables personales como edad, género, inteligencia, motivación o autoestima.

En investigación experimental, a menudo, se denomina **variable dependiente** a la que es consecuencia o depende de variables antecedentes. Una variable que es antecedente o precede a la variable dependiente se denomina **variable independiente, manipulada o experimental**, es decir, la variable que es manipulada o modificada por el investigador para estudiar su efecto sobre una variable dependiente. El efecto de tal manipulación se observa sobre la variable dependiente. Se llama variable dependiente porque su valor depende y varía con el valor de la variable independiente. Suponga que un investigador quiere observar el efecto de la organización temporal del repaso en el rendimiento en estudios sociales. El investigador manipula la temporalización del repaso (inmediato y retardado) y, luego, mide los efectos sobre el rendimiento en los estudios sociales. Después de que se establezca la relación de forma empírica, el investigador puede predecir la variable dependiente a partir de la variable independiente.

En investigación no experimental, como por ejemplo, la investigación descriptiva, correlacional y mediante encuesta, el investigador no manipula las variables directa o activamente. En la investigación descriptiva y en alguna investigación mediante encuesta puede haber sólo una variable de interés. Los estudios que describen el rendimiento en la lectura de los alumnos de segundo curso o las encuestas de las actitudes de los padres hacia la política escolar tienen una sola variable de interés.

En alguna investigación correlacional, la variable antecedente se llama **variable predictora** y la variable a predecir se denomina **variable criterio**. En un estudio que examina la relación de las puntuaciones en el *Scholastic Assessment Test* (SAT)¹ con el éxito en la universidad, la variable predictora son las puntuaciones del SAT y la variable criterio es el éxito en la universidad. En otros estudios correlacionales no hay una variable antecedente obvia. Un ejemplo es un estudio de la relación entre la autoestima y el rendimiento académico. Al investigador no le interesa qué variable precede a otra, sino que, en cambio, quiere conocer la fuerza y dirección de la relación entre las variables. Algunos investigadores emplearán los términos variable dependiente e independiente en investigación correlacional u otro tipo de investigación no experimental cuando quede claro que una variable precede o es antecedente de otra o que las categorías han sido creadas para permitir comparaciones.

Una variable puede ser independiente en un estudio y dependiente en otro; una variable también puede ser predictora o criterio. Que la variable funcione como dependiente o independiente, como predictora o criterio, o como la única variable en un estudio depende del propósito, la lógica y el diseño de la investigación. El enunciado del problema se redacta de modo que indique cuál es la función de las variables en el estudio propuesto.

Observaciones: definiciones operacionales. En la figura 3.1, *Observación(es)* se refiere al método de recogida de datos por medio del cual el investigador razona deductivamente las relaciones con la variable de interés. En investigación cuantitativa, la *observación* se refiere normalmente a una prueba para medir una variable.

En un estudio cuantitativo cada variable debe ser definida operacionalmente y, posteriormente, categorizada, medida o manipulada. Existen dos tipos de definiciones: constitutiva y operacional. Una definición constitutiva, similar a la que encontramos en el diccionario, define un término recurriendo a otros términos. Un diccionario puede definir *ansiedad* como «aprensión o miedo vago» o *inteligencia* como «agudeza mental» y «habilidad de pensar de forma abstracta». Estas definiciones resultan insuficientes para los investigadores. Considere que para *inteligencia*, los sinónimos comunes son parte de la definición constitutiva. Los investigadores emplean definiciones operacionales. Una **definición operacional** asigna significado a una variable especificando las actividades u operaciones necesarias para medir, clasificar o manipular dicha variable. Las definiciones operacionales dicen al investigador y al lector lo que se necesita para responder la pregunta o poner a prueba la hipótesis.

Para mostrar un ejemplo extremo, la hipótesis «la inteligencia se relaciona positivamente con el rendimiento» puede definirse operacionalmente especificando

¹ El SAT es el patrón de las pruebas de admisión comúnmente requeridas en los college y universidades estadounidenses.

cómo medirá el investigador las dos variables. La variable «inteligencia» se mide por medio de las puntuaciones en el test de inteligencia X, o, dicho de otra forma, inteligencia es lo que el test de inteligencia X mide en el estudio. El rendimiento puede ser definido operacionalmente a través de un test estandarizado, un test de rendimiento hecho por el profesor, las notas del alumno u otros métodos de evaluación. Las variables con frecuencia pueden ser definidas operacionalmente de varias formas, siendo algunas más válidas que otras para un problema de investigación. A menudo, la definición operacional de una variable no es tan válida como el investigador desearía. Sin embargo para realizar una investigación, se deben definir operacionalmente todas las variables.

Las preguntas específicas de investigación y las hipótesis pueden operacionalizar las variables. Por ejemplo, la hipótesis «existe una relación positiva entre autoestima y creatividad» puede definirse operacionalmente diciendo «existe una relación positiva entre las puntuaciones de la Escala de Autoestima de Coopersmith y las puntuaciones del *Test Torrance* de Pensamiento Creativo». Tal enunciado indica que el investigador ha elegido estos procedimientos concretos de operacionalización.

Formulación del problema

Un procedimiento eficaz para transformar un tema general en un problema manejable es identificar la población, las variables y la lógica del problema. Supongamos que un supervisor está interesado en diferentes formas de organizar un programa para alumnos de primaria catalogados como de altas capacidades u examinar su efecto en la creatividad de los alumnos. Los programas para niños con altas capacidades han sido organizados de varias formas: como programas especiales en los que los alumnos permanecen juntos para un programa comprensivo; como un programa *pullout* o de desplazamiento, en el que los alumnos asisten regularmente a clase excepto durante dos horas al día en que profesores específicos les imparten clases; y como programas de enriquecimiento en los que los alumnos realizan actividades de enriquecimiento como un complemento a su formación ordinaria. La pregunta se convierte en «¿existen diferencias en la creatividad (variable dependiente) de los alumnos con altas capacidades que participan en un programa especial, de desplazamiento o de enriquecimiento (los tres niveles de la variable independiente)?». Esta pregunta se ha reducido al nivel de identificar la población y las dos variables. La lógica de la pregunta es clara porque se puede identificar la relación entre las variables dependiente e independiente.

Supongamos que se expresa la siguiente pregunta: «¿la integración hace algún bien?». Tal y como está redactada, la pregunta no indica ni la población ni las variables. Un investigador podría decidir que el interés real reside en los

cambios de actitud de los niños sin discapacidad hacia los niños discapacitados. Las actitudes se convierten en variable dependiente y los niños sin discapacidad son una identificación parcial de la población. El investigador decide que, probablemente, el grupo de edad con más posibilidades de experimentar un cambio actitudinal es el de los alumnos de secundaria en lugar de los chicos más pequeños, que pueden estar en proceso de formación de sus actitudes. Los alumnos de secundaria constituyen la población escogida. Además, el investigador decide que el término «integración» es demasiado vago y la variable independiente debería ser la participación de seis semanas en una clase de integración. Ahora, el problema de investigación queda enunciado del siguiente modo: «¿existen diferencias en las actitudes de los alumnos de secundaria (la población) hacia los alumnos discapacitados (la variable dependiente) que participaron en una clase de integración durante seis semanas y la actitud de aquéllos que no lo hicieron (la variable independiente)?». Considere que la pregunta no menciona el cambio, interés principal del estudio. Por lo tanto, la pregunta se puede reescribir como: «¿existen diferencias en las actitudes previas y posteriores hacia los alumnos discapacitados (variable dependiente) entre alumnos de secundaria (la población) que participaron en una clase de integración durante seis semanas y aquéllos que no lo hicieron (variable independiente)?». Ahora la pregunta está bien enfocada y la lógica del problema queda explícita.

Algunos problemas tienen una población y unas variables definidas, pero puede no distinguirse la variable dependiente de la independiente. Si un investigador está interesado en la autoestima y el rendimiento académico, ¿cuál es la variable dependiente? ¿El problema es investigar si el rendimiento se relaciona con la autoestima positiva (la variable dependiente) o si la autoestima se relaciona positivamente con el rendimiento (la variable dependiente)? En este caso la dificultad reside en la naturaleza de las variables. En estos problemas, se manejan variables categóricas y continuas, pero no variables experimentales o manipuladas. El problema se puede enunciar de la siguiente manera «entre los alumnos de sexto curso, ¿existen diferencias en su rendimiento académico entre los que tienen baja autoestima y aquéllos cuya autoestima es alta?». La población está constituida por los alumnos de sexto curso y las variables son la autoestima y el rendimiento académico. No hay ninguna variable manipulada ya que los sujetos ya tienen esos atributos, la autoestima y el rendimiento académico. Tampoco se clasifica ninguna variable como dependiente o independiente porque no se puede determinar cuál antecede a la otra.

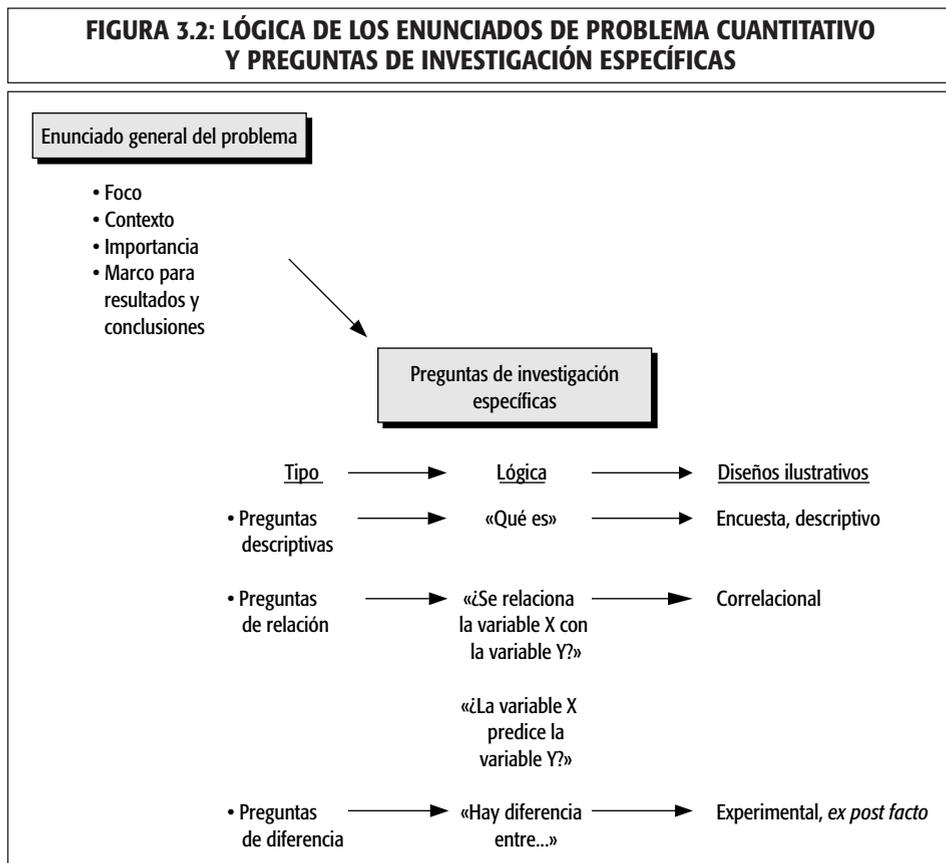
A través de la identificación del constructo principal, las variables y la población, el investigador aclara el foco y la lógica del problema. Este proceso no es fácil y, como se ha mencionado anteriormente, no puede realizarse sin un trabajo previo. Leer la bibliografía, participar en tormentas de ideas con otros y hablar con investigadores experimentados puede ayudar a definir un problema. Ya con

las ideas claras, el investigador puede redactar un planteamiento formal del problema. El planteamiento formal del problema puede ser redactado como enunciados del propósito de la investigación, como preguntas de investigación específicas o como hipótesis de investigación, dependiendo del propósito del estudio y del diseño seleccionado. A continuación se explican cada una de estas formas de definir problemas específicos de investigación.

Preguntas de investigación específicas

En los estudios cuantitativos el problema de investigación puede ser enunciado en forma de pregunta. Se prefiere este formato porque es simple y directo. Psicológicamente, orienta al investigador sobre la tarea siguiente: desarrollar un diseño para contestar la pregunta. Las preguntas de investigación pueden ser descriptivas, de relación o de diferencia. Cada tipo de pregunta implica un diseño diferente (ver figura 3.2).

FIGURA 3.2: LÓGICA DE LOS ENUNCIADOS DE PROBLEMA CUANTITATIVO Y PREGUNTAS DE INVESTIGACIÓN ESPECÍFICAS



Preguntas descriptivas de investigación. Normalmente preguntan «qué es» e implican diseños de investigación descriptivos y de encuesta. Sin embargo, estos términos no siempre se emplean en la redacción de las preguntas de investigación. Por ejemplo, una pregunta de investigación puede ser: «¿cuál es el nivel de ejecución en el test de Iowa de habilidades básicas de nuestros alumnos de cuarto curso?». La investigación mediante encuesta estudia, a menudo, las percepciones de los grupos acerca de una práctica, como por ejemplo: «¿cuál es la opinión de los directores de un programa?», «¿cuáles son las actitudes que muestran nuestros alumnos hacia los niños de integración?», «¿cuál de las rutas alternativas del autobús prefieren los padres de nuestros alumnos?» o «¿cuáles son las necesidades de instrucción más importantes según la plantilla?».

Preguntas de relación. Las preguntas de relación se plantean «cuál es la relación entre dos o más variables» e implican un diseño correlacional. Esto no significa que siempre aparezcan en el enunciado las palabras exactas: «¿cuál es la relación entre la variable A y la variable B?». Por poner un ejemplo, «¿la autoestima se relaciona con el rendimiento?» interroga acerca de la relación entre una variable (autoestima) y otra variable (rendimiento). Los estudios que determinan los mejores mecanismos de predicción de una variable, como por ejemplo, previsiones de éxito en la universidad, implican preguntas de relación entre las posibles variables de previsión, tales como la nota media en secundaria, las influencias y la participación en actividades extracurriculares, y la variable dependiente: éxito en la universidad. El extracto 3.2 ilustra un enunciado de problema que implica una pregunta de relación entre mecanismos de predicción a largo plazo y el ajuste social de los niños. El enunciado del problema sugiere el diseño del estudio.

Preguntas de diferencia. Las preguntas de diferencia, normalmente, cuestionan si «existen diferencias entre dos grupo, dos o más tratamientos o dos conjuntos de datos (medidas)». Se utilizan cuando el estudio compara dos o más observaciones. Enunciar la pregunta como «¿existen diferencias?» en lugar de «¿existe relación?» entre dos o más observaciones evidencia la lógica subyacente del estudio. Preguntas como por ejemplo: «¿existen diferencia entre las puntuaciones del pretest y del postest?» son más útiles que aquéllas redactadas como «¿existe relación entre las puntuaciones del pretest y del postest?»

Los diseños experimentales y *ex post facto* emplean preguntas de diferencia. El extracto 3.3 ilustra las preguntas de investigación que requieren comparación de dos tratamientos (instrucción tradicional con instrucción por medio de juegos de simulación) en tres variables dependientes diferentes.

Si los investigadores creen firmemente que, además de predecir la diferencia entre dos o más observaciones, pueden predecir la dirección en la que reside la diferencia, entonces la dirección queda definida en la pregunta de investigación. La pregunta: «¿existen diferencias entre las puntuaciones del pretest y del

EXTRACTO 3.2

PREGUNTAS DE RELACIÓN

Los objetivos del presente estudio eran: (a) identificar los correlatos o mecanismos de predicción a largo plazo del ajuste social y (b) determinar la correlación múltiple entre el mejor conjunto de esos mecanismos de predicción y el ajuste social de los niños después de 6 o 9 años. Los sujetos eran niños que fueron evaluados primero en tercer o sexto curso y cuyo ajuste social se volvió a evaluar 6 o 9 años después.

Fuente: Feldhusen, J. F., Roeser, T. D. y Thurston, J. R. (1977), «Prediction of social adjustment over a period of 6 or 9 years», *Journal of Special Education*, 11 (1), 31.

postest?» puede enunciarse como «¿existe un mayor nivel de comprensión lectora en el postest comparado con el pretest?»

Las preguntas de investigación no son preguntas estadísticas enunciadas para el análisis de datos. Las preguntas estadísticas pueden enunciarse como: «¿hay diferencias estadísticamente significativas entre las observaciones A y B?» o «¿hay relación estadísticamente significativa entre las variables A y B?». En una investigación, las preguntas estadísticas se enuncian en la sección de metodología. Las preguntas de investigación se definen en la introducción de un estudio y sugieren el diseño posterior.

Hipótesis de investigación

Una **hipótesis de investigación** es un enunciado provisional de la relación esperada entre dos o más variables. El enunciado describe, en otras palabras, los

EXTRACTO 3.3

PREGUNTAS DE DIFERENCIA

PI 1: ¿Existen diferencias significativas en el aprendizaje afectivo de los alumnos entre la instrucción tradicional en el aula y la enseñanza mediante simulación?

PI 2: ¿Existen diferencias significativas en el aprendizaje cognitivo de los alumnos entre la enseñanza tradicional en el aula y la enseñanza mediante simulación?

PI 3: ¿Existen diferencias significativas en la motivación de los alumnos entre la enseñanza tradicional en el aula y la enseñanza mediante simulación?

Fuente: Garard, D. L., Hunt, S. K., y Paynton, S. T. (1998), «Alternative to tradicional instruction: Using games and simulation to increase student learning and motivation». (p. 38). *Communication Research Reports*, 15 (1), 36-44.

resultados previstos. Los enunciados del problema y las hipótesis de investigación son similares en su esencia, excepto en que las hipótesis de investigación son enunciados declarativos, más específicos que los enunciados del problema, claramente comprobables e indicativos de los resultados esperados. Para el problema de investigación: «¿existe relación entre repaso y recuerdo?», la hipótesis de investigación podría ser «existe una relación positiva entre repaso y recuerdo». La comprobación empírica es factible en tanto que cada variable puede ser manipulada, clasificada o medida. Si no se puede manipular, clasificar o medir una variable, no existe un método de análisis cuantitativo para la comprobación de la hipótesis.

Tres ejemplos de hipótesis de investigación son: «cuanto mayores son las diferencias percibidas entre adultos y adolescentes como aprendices, mayores son las diferencias en los comportamientos de aprendizaje»; «los alumnos con bajo nivel de lectura obtendrán un mayor rendimiento lector en un programa de intervención en la lectura que alumnos comparables en un curso de literatura inglesa»; y «el estilo de dirección democrática produce más satisfacción en el profesorado que el estilo autoritario». Aunque la palabra *relación* no se usa en todas las hipótesis, las expresiones de relación como por ejemplo *obtendrán un mayor rendimiento, produce, en función de y efectos* conectan las variables.

Una hipótesis implica una lógica si A—entonces B. La mayoría de las hipótesis puede ajustarse a esta estructura para indicar la relación entre variables. En cada uno de los tres ejemplos anteriores, la hipótesis tiene una variable dependiente y una independiente: si se perciben diferencias, entonces mayores son las diferencias en el comportamiento del maestro; si hay un programa de intervención, entonces mayor comprensión lectora; y si el estilo de dirección es democrático, entonces mayor satisfacción del profesorado. La lógica es similar para hipótesis con más de dos variables.

Una hipótesis es una explicación, en forma de suposición, de un fenómeno que se acepta o se rechaza por una evidencia experimental. En los ejemplos anteriores, las diferencias percibidas entre adultos y adolescentes como aprendices explican las diferencias en los comportamientos docentes; un programa de intervención en la lectura explica un mayor rendimiento y un estilo de dirección democrática da cuenta de la satisfacción del profesorado.

Las hipótesis son especialmente importantes en la investigación correlacional y experimental que estudian relaciones entre variables. Una hipótesis debe cumplir varios patrones para ser útil en investigación:

1. **La hipótesis debe definir la relación esperada o la diferencia entre dos o más variables.** Un enunciado como por ejemplo: «si *feedback* del profesor, entonces rendimiento en ciencias» implica una relación pero no es una hipótesis. La hipótesis direccional debería enunciarse «el *feedback* del profesor está positivamente relacionado con el rendimiento en ciencias» o «existe una relación positiva entre el *feedback* del profesor y el rendimiento del alumno en ciencias».

En investigación experimental, en la que se administra un tratamiento a un grupo de sujetos pero no a otro grupo, los investigadores deberían intuir diferencias direccionales; por ejemplo: «los alumnos de quinto curso que han recibido enseñanza por ordenador obtendrán un mayor rendimiento en matemáticas que los alumnos comparables que no la recibieron».

2. **Una hipótesis debe ser comprobable.** Una hipótesis comprobable debe ser verificable; uno puede esbozar conclusiones desde observaciones empíricas que indiquen si se puede apoyar o no una relación. El investigador puede determinar si las hipotéticas consecuencias ocurrieron o no. Para ser comprobable, una hipótesis debe incluir variables relacionadas que puedan medirse o clasificarse por medio de algún procedimiento objetivo. Por ejemplo, puesto que se puede clasificar a los alumnos de primer curso por su asistencia o no a preescolar, una hipótesis podría enunciar: «los niños que asisten a preescolar tendrán mayor puntuación en una escala de madurez social que los niños que no asisten a preescolar». Si una variable no puede ser medida o clasificada, no existe un método para hacer las comparaciones estadísticas necesarias.

3. **Una hipótesis debería ofrecer una explicación provisional basada en una teoría o investigación previa.** Una hipótesis bien fundamentada indica que hay suficiente investigación o teoría como para ser considerada lo suficientemente significativa para ser puesta a prueba. Una hipótesis de investigación se define después de la revisión de la bibliografía; esto significa que el investigador conoce los trabajos previos. Una hipótesis normalmente no es disconforme con la evidencia anterior predominante; pero, si es puesta a prueba, puede aumentar nuestro conocimiento del problema de investigación. En ciencias sociales, si la hipótesis está teóricamente fundamentada es posible que los resultados del estudio enriquezcan la teoría objeto de estudio. Sin embargo, en muchas áreas de la educación la evidencia concluyente es escasa y sólo parte de la investigación educativa puede servir como base para la hipótesis de investigación.

4. **Una hipótesis debería ser concisa y lúcida.** Una hipótesis, en su forma más sencilla, debería tener coherencia lógica y un orden evidente en su organización. Los enunciados breves ayudan tanto al lector como al investigador en la interpretación de los resultados. Una regla general es enunciar sólo una relación por hipótesis. Aunque un investigador puede tener una hipótesis general amplia, es mejor en aras de la claridad, redactar el enunciado como hipótesis más específicas.

FORMULACIÓN DEL PROBLEMA EN LA INVESTIGACIÓN CUALITATIVA

La formulación del problema en la investigación cualitativa comienza con la selección de un tema general y una modalidad de investigación (interactiva

o no interactiva). El tema y la metodología están interrelacionados y se seleccionan casi simultáneamente mejor que de forma separada. Por ejemplo, una inicial decisión de investigación es si examinar sucesos actuales o pasados. Un estudio de fenómenos actuales requiere que el investigador tenga acceso a una situación o a un grupo de personas que posean alguna experiencia social compartida, como por ejemplo, trabajar en el mismo sistema escolar, participar en un proyecto especial o en una clase y asuntos similares. Un estudio de hechos pasados requiere colecciones de archivos de documentos originales que estén disponibles y accesibles al investigador. Estas consideraciones comienzan a dar forma y a influir en la selección de un tema general.

Los investigadores cualitativos comienzan reduciendo un tema general a otro más concreto. Supongamos que alguien está interesado en la enseñanza de la lectura y la escritura. Existen diversas orientaciones teóricas para la lectoescritura, como por ejemplo, la decodificación, las habilidades y destrezas y el método global. El investigador está interesado en un enfoque global y sus observaciones personales y experiencia sugieren que un aula de método global opera de manera diferente a una clase de *fonología* o de *destrezas*. Por ejemplo, no habría abecedarios, ni conjuntos de textos de lectura de legibilidad controlada, ni tareas escritas. En su lugar, la escritura de los niños estaría integrada en un programa de lectura para niños con «textos globales significativos» como materiales didácticos en lugar de «palabras aisladas, sonidos, o “historias” con vocabulario controlado» (Edelsky, Draper, y Smith, 1983, p. 259). El investigador empieza a preguntarse cómo consigue un maestro que un niño lea y escriba, de forma correcta, para cumplir las expectativas en esta inusual situación.

Otro ejemplo, supongamos que un investigador ha observado que la evaluación por medio de tests en los colegios públicos se ha incrementado en la última década y ha escuchado críticas de que los profesores, directores y orientadores emplean la aplicación de tests para segregar a los niños y promover el control social a través de los «itinerarios» y el agrupamiento por capacidades. Sin embargo, el investigador está interesado en saber si el pasado nos posibilita comprender cómo se empleó originalmente la evaluación por medio de tests: en el pasado, ¿se aceptaron los tests como un instrumento científico para la separación de alumnos por su inteligencia o resultaron ser medidas de evaluación poco fiables para los niños negros y mexicanos? El asunto de los tests en los colegios públicos se convierte, entonces, en un tema más preciso de evaluación de inteligencia y de sus propiedades de medida cuando se aplica en los colegios públicos para el agrupamiento o «la separación por itinerarios» de los niños.

La mayor parte del interés por la investigación cualitativa procede de las experiencias personales y de un fuerte empeño en un tema desarrollado a partir de las oportunidades de la biografía y de la historia personal. Estas oportunidades proporcionan al investigador acceso físico y/o psicológico a los escenarios sociales

presentes o pasados (Riemer, 1977). Tal acceso se convierte en el punto de partida para la investigación cualitativa significativa sólo cuando va acompañado por algún grado de interés o de compromiso. Ejemplos de una biografía son un trabajo, un cambio en una relación íntima o una actividad de la que se disfruta. La historia personal se refiere a alguna experiencia anterior, como por ejemplo, un investigador que adquiere interés sobre cómo las mujeres trabajadoras que están casadas organizan sus cambiantes vidas debido a las variaciones que suceden en su propia vida. En otras palabras, los problemas de investigación descansan en muchas situaciones personales y experiencias y lecturas generales, que sólo necesitan ser reconocidas como problemas de investigación potenciales. El conocimiento, la comprensión y la reflexión acerca de los modos de investigación cualitativa permiten al investigador seleccionar la lógica para el problema.

La lógica inductiva de los registros de campo cualitativos, las descripciones y las abstracciones

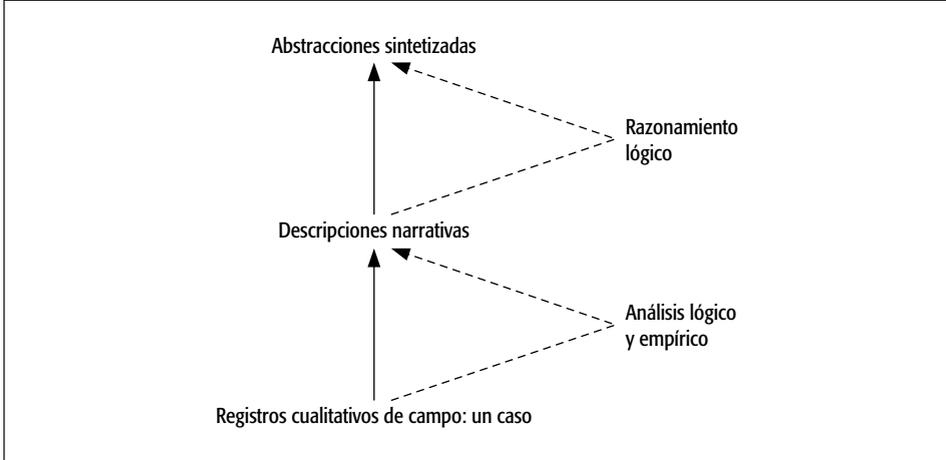
La investigación cualitativa, al contrario que la investigación cuantitativa, emplea principalmente el *razonamiento inductivo*. El «problema» se enuncia con mayor claridad después de la recogida y el análisis preliminar de los datos. El investigador obtiene muchos registros de campo de una situación presente o pasada, con los que elabora descripciones detalladas de las percepciones de la gente y de las realidades sociales, generando abstracciones sintetizadas de estas descripciones para explicar el fenómeno. El razonamiento inductivo permite *explorar* y *descubrir* con un diseño de investigación emergente en lugar de comprobar deducciones de una teoría en un diseño predeterminado. El problema de investigación se reformula, normalmente, durante la recogida de datos de tal manera que los datos representan fielmente la realidad de las experiencias sociales compartidas.

La formulación del problema comienza con la selección de un caso particular para un estudio en profundidad. El proceso inductivo se representa esquemáticamente en la figura 3.3

Observe que la dirección de las flechas es de abajo hacia arriba para ilustrar el razonamiento inductivo. Vamos a explicar cada término de la figura 3.3, empezando por los registros cualitativos de campo de un caso seleccionado.

Registros cualitativos de campo: un caso. Los investigadores seleccionan un caso específico para su estudio en profundidad. Un caso es una situación determinada elegida por el investigador en la que algún fenómeno será descrito a partir de las percepciones de los participantes. El investigador también selecciona una metodología cualitativa principal (observación participante, entrevista en profundidad o análisis de documentos). Por ejemplo, en un estudio de «Hookin» *Em In at the Start of School in «Whole Language» Classroom* (Edelsky, Draper y Smith, 1983), el caso seleccionado fue una clase de sexto curso cuyo profesor

FIGURA 3.3: LÓGICA INDUCTIVA EN INVESTIGACIÓN CUALITATIVA: OBSERVACIONES, DESCRIPCIONES Y CONCEPTOS



poseía una visión global de la capacidad de leer y escribir. El caso elegido para un estudio sobre los orígenes del extracurrículo de instituto sucedió en Chicago, 1880-1915 (Gutowski, 1988), en el que se investigaron a partir de documentos históricos del sistema educativo de Chicago, periódicos escolares, anuarios de los alumnos, diarios de la época y el *Chicago Tribune*. El caso seleccionado limita el problema de investigación a un contexto particular, en nuestro ejemplo, una clase de sexto curso o un sistema escolar en la década de 1930 o el extracurricular en un sistema educativo desde 1880 hasta 1915.

Los investigadores seleccionan un caso particular, en lugar de variables, a través del que logran la comprensión de un suceso más amplio. En cada uno de los ejemplos citados arriba, el suceso más amplio era la enseñanza de la lectoescritura y el programa extracurricular del instituto.

Los registros cualitativos de campo (datos) obtenidos a lo largo del tiempo, se registran como notas de campo de observación participante, entrevistas en profundidad o notas del investigador de documentos históricos. Cada nota de campo, transcripción o documento histórico contiene la fecha y el contexto en que se produjo, como por ejemplo, las circunstancias sociales, la situación y los participantes. Los datos de campo iniciales pueden llevar a la recogida de datos de otras personas, otros escenarios u otros archivos con los que el investigador descubre más aspectos sobre el caso seleccionado. En otras palabras, el término «caso» no hace referencia a una persona, una colección de archivos o una localidad, sino a la *situación social* objeto de estudio. Las estrategias de recogida de datos se ajustan para obtener una visión global del fenómeno y, luego, estudiar ciertos aspectos en profundidad. En el extracto 3.4 el caso seleccionado fue un grupo de profesoras

en un lugar concreto, Hamilton, Ohio. El investigador eligió examinar un fenómeno en profundidad con un diseño de estudio de caso en lugar de con un diseño experimental o no experimental.

Las descripciones narrativas. Los investigadores utilizan los datos para construir descripciones narrativas, narraciones detalladas de personas, de incidentes y de procesos. La narración descriptiva queda terminada por completo después de la recogida de datos debido a la orientación hacia el descubrimiento de la investigación. Para generar, de forma inductiva, una narración descriptiva debe contarse con ciertos tipos de datos en las notas de campo, transcripciones o notas de documentos históricos/jurídicos.

Las narraciones descriptivas, algunas veces denominadas descripción «rica» o «densa», «con todo tipo de detalles», contienen por lo menos cuatro elementos: personas, incidentes, lenguaje de los participantes y «significados» de los participantes. (1) Se describen los participantes como individuos que tienen historias personales diferentes y presentan características físicas, emocionales e intelectuales distintas en situaciones variadas. (2) Los incidentes constituyen la narración de la situación social, es lo más parecido a contar una historia. (3) Las descripciones enfatizan el lenguaje de los participantes, no el del investigador o el de la ciencia social. Se anotan los nombres de los participantes en los sucesos, lugares, objetos, acontecimientos especiales y procesos. El lenguaje se refiere a las muchas formas de comunicación como por ejemplo, la expresión verbal y no verbal, los esquemas, los dibujos, los símbolos y elementos similares. (4) Las descripciones subrayan los «significados» de los participantes. Estos son los puntos de vista de la gente acerca de la realidad o de cómo perciben su mundo. Los significados de los participantes se transmiten cuando una persona define «por qué» ha sucedido algo.

EXTRACTO 3.4

CASO SELECCIONADO PARA ENTREVISTAS EN PROFUNDIDAD

Si falló el sindicalismo es porque la gente normal toma sus decisiones con respecto a sus propios mundos. La atención prestada a las fuerzas mayores de la historia proporciona un marco de comprensión; pero, sin una descripción de los detalles más sutiles de las realidades subjetivas de los participantes, fallamos completamente al entender la dinámica de la historia ... [la cual] ... a menudo involucraba a las mujeres que vivían en sus mundos subjetivos, pero igualmente reales.

Este artículo es una historia oral. Su objetivo es describir la realidad subjetiva compartida de un grupo de maestras durante los años treinta y comparar esa realidad con algunas apreciaciones que han realizado los historiadores acerca del sindicalismo. Está basado, principalmente, en entrevistas orales e influido por la etnohistoria, el artículo pre-

senta una definición grupal de los colegios y la vida en una pequeña ciudad del medio oeste durante la Gran Depresión. Desde este punto de vista, nos proporciona una perspectiva antropológica al estudio histórico. Al intentar descubrir las definiciones culturales de los participantes en una situación histórica, estudia el lado subjetivo de la historia. Este artículo está menos interesado en describir cómo fueron realmente los hechos que en explorar la forma en que fueron percibidos por los participantes. Como estudio de caso, no reclama la posesión de una verdad generalizable.

Fuente: Quantz, Richard A. «The Complex Visions of Female Teachers and the Failure of Unionization in the 1930s: An Oral History», *History of Education Quarterly*, vol. 25, No. 4, 1985. Reimpreso con permiso del autor.

Las abstracciones sintetizadas. Los investigadores conciben, de forma inductiva, abstracciones a partir de los datos. Las **abstracciones sintetizadas** son generalizaciones y explicaciones resumidas de los principales hallazgos de un estudio. Estas abstracciones sintetizadas pueden tomar formatos diferentes como por ejemplo, una lista de temas narrativos, «lecciones aprendidas», lo esencial de una experiencia compartida, la descripción de un concepto, afirmaciones o enunciados de propuestas. El investigador construye una imagen que toma forma cuando recoge los datos y examina las partes. Los significados sutiles de un fenómeno pueden ser comprendidos más claramente por los lectores. Por ejemplo, un estudio explicaba los significados del fracaso del sindicato de profesores mediante las descripciones de los profesores de su relación con los colegios. Los profesores referían esta relación en términos de «cuatro metáforas: la figura subordinada de la autoridad, el colegio como familia, la vocación natural de la mujer y el ego dual» (Quantz, 1985, p. 442). Sólo un profesor no empleó estas metáforas.

Reformulaciones del problema

Los problemas de investigación cualitativa se reformulan varias veces *después* de que el investigador haya comenzado la recogida de datos. El problema de investigación se define inicialmente al planificar el estudio, se reformula durante el comienzo de la recogida de datos y, de nuevo si es necesario, se reformula a lo largo de la recogida de datos. La reformulación continua del problema de investigación refleja un diseño emergente. Las reformulaciones de un problema de investigación reflejan las modificaciones en las estrategias de recogida de datos para adquirir la «totalidad» de los fenómenos y, después, el estudio de algún aspecto en profundidad. El problema específico de investigación evoluciona en consecuencia y se abrevia hacia el final de la recogida de datos. La versión abreviada del problema de investigación, en la mayoría de las publicaciones, no coincide con el enunciado de problema que inició la investigación.

Problemas previstos. Los investigadores cualitativos comienzan con **problemas previstos**, problemas de investigación anticipados que serán reformulados durante la recogida de datos (Malinowski, 1922). El planteamiento de los problemas pronosticados depende, en cierta medida, del conocimiento previo de los sucesos y de los procesos de una situación, de la gente entrevistada o de la colección de documentos de archivo. De este modo, los problemas pronosticados no derivan directamente de una revisión exhaustiva de la bibliografía, sino más bien de las experiencias iniciales que un investigador adquiere planificando el estudio. Sin embargo, una búsqueda bibliográfica preliminar ayuda al investigador en la redacción de los problemas pronosticados.

Los problemas pronosticados se redactan, frecuentemente, en forma de preguntas generales, preguntas centradas en el «qué», el «cómo» y el «porqué» de la situación. El «qué» se refiere a quién, cuándo, dónde y cuáles situaciones sociales tuvieron lugar. El «cómo» hace referencia a los procesos a analizar y qué influyó sobre éstos. El «porqué» se refiere a los «significados» del participante, sus explicaciones de los hechos y las situaciones sociales observadas. Cada una de estas preguntas es amplia, de forma deliberada, debido a la lógica de descubrimiento de un diseño emergente.

Los problemas pronosticados se formulan, con frecuencia, durante las primeras fases de la recogida de datos. Los etnógrafos pueden pasar de las observaciones a las entrevistas; los entrevistadores pueden ajustar las preguntas a las descripciones de temas no planificados originalmente; los investigadores históricos examinan documentos, a menudo, para decidir lo que es importante recoger de los registros escritos.

Un estudio etnográfico de una aula de *método global del lenguaje* ilustra este proceso (ver extracto 3.5). El problema previsto basado en el conocimiento preliminar fue «¿cómo consigue este profesor que el niño cumpla sus inusitadas expectativas?, ¿cuáles son aquí las normas para leer y escribir?, ¿cómo consigue el profesor que los alumnos deseen escribir?, ¿cómo están establecidos ciertos procedimientos (por ejemplo, el cuaderno de escritura)?, ¿cuáles son las relaciones observables entre profesor y alumno?» (Edelsky, Draper y Smith, 1983, p. 261-262). Aunque los investigadores esperaban descubrir un cambio gradual, era evidente que los alumnos ya estaban empezando a cumplir los deseos del profesor durante el primer día de curso. Los investigadores modificaron sus planes y entrevistaron a los niños el segundo día de clase. Esto obligó a reformular el problema: «¿qué está ocurriendo aquí?». La reformulación final y el planteamiento resumido del problema se convirtió en «¿cómo consigue el profesor que la vida en la clase se convierta en lo que él quiere, es decir, cómo ocurre esto en una clase con un profesor eficaz que aplica un método global de aprendizaje de la lectoescritura» (Edelsky, Draper y Smith, 1983, p. 260).

Enunciado resumido del problema. Los enunciados resumidos del problema pueden ser escritos en cualquier momento, durante o después de la recogida de datos. Normalmente, el enunciado resumido del problema coincide con la pregunta principal de investigación, que es el núcleo de todo el informe. El título, la revisión bibliográfica y la discusión emplean, a menudo, términos abstractos de investigadores cualitativos; sin embargo, el problema de investigación puede ser o no redactado en términos descriptivos. Por ejemplo, en *Individualism and Community: Ritual Discourse in a Parochial High School* («Individualismo y comunidad: discurso ritual en un instituto parroquial»), el problema de investigación es «tales lazos permanentes (con la escuela *St. Anne's*) no se constituyen sólo con buenas intenciones; deben ser reafirmados a través de repetidas experiencias concretas que se consideran ‘buenas’ y ‘dignas de consideración’». ¿Qué pasó con la relación con *Sainte Anne's* cuando la escuela restringió la libertad de movimiento de los alumnos mediante un campus cerrado, limitó su elección de cursos y añadió cargas financieras a las familias de muchos alumnos? (Lesko, 1986, p. 26). El título «Individualismo y comunidad...» hace referencia a conceptos procedentes de la antropología y de la sociología.

Objetivos y preguntas de investigación cualitativa

Los estudios cualitativos contienen enunciados de objetivos de investigación y de preguntas que conllevan la lógica inductiva ante el problema. El planteamiento del objetivo de la investigación es la versión final resumida del planteamiento inicial del problema. Las preguntas específicas de investigación pueden estar enunciadas o ser implícitas. Los enunciados de los objetivos implican la modalidad cualitativa elegida: formas de investigación interactiva o no interactiva.

EXTRACTO 3.5 REFORMULACIONES DEL PROBLEMA Y DISEÑO EMERGENTE

Lo que afrontaban los alumnos de sexto curso en nuestra investigación, sin embargo, eran las expectativas de uno de los profesores de la minoría que trabajaba a partir de un método global de la enseñanza de la lectoescritura. Karen Smith (a partir de ahora KS), profesora de sexto curso y una de las autoras, no utilizó libros de texto, ni cuadernos, ni cartillas...

El fenómeno que intentábamos explicar fue cómo, al principio del año, los profesores «convencen» a los niños... para que la vida en el aula se convierta en lo que ellos desean que sea. Sin embargo, no estábamos sólo interesados en cómo

**Planteamiento
del problema
resumido final**

sucede esto en cualquier aula, sino también en cómo ocurre con un profesor eficaz que tiene un punto de vista global de la enseñanza de la lectoescritura...

Nuestra recogida de datos principal se llevó a cabo a través de la observación participante de la interacción entre profesor y alumno a lo largo de todo el día y durante todos los días de las dos primeras semanas y tres días por semana de las tres semanas siguientes. Periódicamente, se realizaron grabaciones de audio y de vídeo que fueron utilizadas para modificar el foco de interés de observaciones posteriores. Recogimos notas de campo durante las observaciones del aula y el visionado de las grabaciones. Se entrevistó a los alumnos durante la primera y la tercera semana. Entrevistamos a KS antes del comienzo del curso y durante la cuarta semana. Volvimos en diciembre y, de nuevo, en enero para verificar si los tipos de interacción observados en septiembre todavía se mantenían.

Ningún investigador puede observarlo todo. Las decisiones sobre qué estudiar, estaban basadas, en parte, en el conocimiento previo de los investigadores. Además del conocimiento de los resultados de investigación, empezamos con el considerable conocimiento previo de esta maestra. Tanto la recogida como el análisis de datos se basaron en este conocimiento previo; uno no puede desconocer lo que sabe.

Empezamos con la pregunta: ¿cómo consigue esta maestra que los niños puedan alcanzar sus inusitadas expectativas? Nuestras observaciones iniciales se guiaron por una variedad de preguntas: ¿cuáles son aquí las reglas para leer y escribir?, ¿cómo consigue el profesor que los alumnos deseen escribir?, ¿cómo están establecidos ciertos procedimientos (por ejemplo, el libro y cuaderno de escritura)?, ¿qué relaciones se evidencian entre el profesor y el alumno?

Esperábamos ver un cambio gradual durante las primeras semanas de colegio, con algunos alumnos que dudaban de aceptar las exigencias poco comunes de KS, otros adoptando el nuevo concepto de colegio y de lectura casi de inmediato y otros tomándose un tiempo considerable para su adaptación. Sin embargo, durante el primer día de colegio fue evidente que los alumnos ya se habían convertido en lo que el profesor quería que fueran. Esa misma tarde hacían la limpieza sin que se les pidiera, se ayudaban unos a otros y asumían responsabilidad para tomar decisiones y realizar las tareas asignadas. Teníamos aquí unos niños, en un ambiente relativamente nuevo, actuando casi inmediatamente como de forma «innata».

Estrategias de recogida de datos

Conocimiento preliminar a partir del planteamiento de la investigación

Problemas previstos

Expectativas iniciales y primer día de recogida de datos

Nos dimos cuenta rápidamente de que habíamos establecido una suposición errónea con respecto a una adaptación gradual. Nuestro plan original contemplaba las entrevistas a los alumnos durante la primera semana de colegio, pero las evidencias de que nuestro objeto de interés se estaba precipitando nos obligaron a entrevistar a algunos niños el segundo día de colegio. Estos niños se habían dado cuenta de que su clase era diferente; hablaban de la ausencia del ejercicio de habilidades de las clases tradicionales, como el deletreo, los libros de texto y aprendizaje de nombres. Pero no podían expresar aquello que KS esperaba o cómo supieron lo que tenían que hacer...

Después del segundo día de colegio, empezamos a buscar reglas no definidas (comprensión tácita) como explicación a cómo los alumnos sabían cuándo tenían que hacer exactamente lo que la profesora decía o cuándo podían seguir una idea más general y actuar con sensatez. A nuestra pregunta inicial (¿cómo esta profesora, con su teoría de la lectoescritura, conseguía que los alumnos cumplieran sus demandas?), ahora habíamos añadido una segunda: ¿qué está pasando aquí? El proceso de plantearse (conjugarse) ambas cuestiones era cíclico, requiriendo un cambio de enfoque desde lo extenso hasta lo concreto y viceversa. Al ver las cintas de video, era evidente que para dar respuesta a ambas preguntas resultaba necesario tener en cuenta lo que estábamos observando; una profesora que conseguía las cosas a su modo, pero que ciertamente no coincidía con la descripción de un profesor eficaz al comienzo del curso, según se desprende de la bibliografía sobre colegios eficaces.

Fuente: "Hookin' 'Em In at the Start of School in 'Whole Language' Classroom" por Carole Edelsky, Kelly Draper y Karen Smith. Reimpreso con permiso de la *Association from Anthropology & Education Quarterly* 14:4, 1983. No se permite la reproducción posterior.

Estrategias de recogida de datos cambiadas y resumen de datos

Primera reformulación del problema de investigación

Enunciados y preguntas de problemas etnográficos. Las prácticas de la investigación cualitativa de etnografía, fenomenología, estudios de caso, teoría fundamentada y estudios críticos se centran en fenómenos actuales de los que se pueden obtener datos por medio de la interacción con los participantes en una situación social seleccionada. El extracto 3.6 ilustra un enunciado de problema etnográfico y las preguntas de investigación en un estudio de aceptación entre hermanos con graves discapacidades. La lógica del planteamiento del problema implica el diseño de investigación.

Enunciados y preguntas de problemas históricos. Los problemas históricos de investigación están centrados en acontecimientos pasados y requieren acceso a documentos en archivos históricos. Los problemas de investigación, a menudo, se generan cuando se ponen a disposición de los estudiosos los documentos de los archivos. El extracto 3.7 ilustra un estudio histórico centrado en una tesis doctoral del año 1939 realizada por Horace Mann Bond, que llegó a ser administrador, profesor y rector de varias universidades para personas de color durante su

EXTRACTO 3.6

PLANTEAMIENTO DE PROBLEMA Y PREGUNTAS DE INVESTIGACIÓN ETNOGRÁFICA

Las relaciones sociales de las personas con discapacidad han sido una preocupación central del campo educativo durante muchos años. Igualmente ha sucedido con las relaciones entre hermanos. Este artículo presenta datos conjuntos de ambos tipos de relaciones, en un intento de ampliar nuestro conocimiento acerca de lo que se necesita para apoyar las relaciones entre jóvenes con discapacidades significativas y sus iguales sin discapacidad. Comentaremos específicamente los resultados de un estudio de caso de las relaciones entre hermanos de Raúl, un joven con síndrome de Down...

Nos centramos en las siguientes preguntas:

1. ¿Cuáles son los tipos predominantes de interacción entre Raúl y sus hermanos?
2. ¿Cómo participa Raúl en esas interacciones?
3. ¿Cómo se comparan las interacciones entre hermanos observadas con los resultados de las interacciones de Raúl con sus iguales sin discapacidad en el colegio?
4. ¿Qué lecciones podemos aprender de estas relaciones entre hermanos para facilitar la inserción de un estudiante con discapacidad?

Fuente: Harry, B., Day, M. y Quist, F. (1998), «He Can't Really Play: An ethnographic study of sibling acceptance and interaction» (p. 289, 291), *JASH: Journal of the Association for Persons with Severe Handicaps*, 23 (4) 289-299.

EXTRACTO 3.7

PLANTEAMIENTO DE PROBLEMA Y PREGUNTAS DE INVESTIGACIÓN HISTÓRICA

La reciente puesta a disposición de los investigadores de los documentos de Horace Mann Bond, junto con la publicación de un artículo sobre los inicios de la carrera de Bond que maneja estos documentos, debería dar arranque a un interés renovado por el trabajo de este notable profesor, investigador y administrador educativo.... Escribió críticas notables sobre la teoría de evaluación psicológica en varias etapas de su larga carrera, realizó investigación histórica para los demandantes en el juicio por segregación de

Brown contra el colegio Topeka y viajó extensamente por África y estudió sus civilizaciones en las últimas etapas de su carrera.

Una de las mayores contribuciones de Bond a la educación es el tema de este ensayo: su tesis doctoral que fue publicada en 1939 bajo el título *Educación negra en Alabama: un estudio en algodón y metal*. Los méritos del libro de Bond se hicieron notar casi inmediatamente después de su publicación...

La importancia que sobresale del trabajo de Bond se refleja en numerosas citas en recientes trabajos históricos, historia de los afroamericanos e historia de la Educación... Este ensayo... examina brevemente las circunstancias personales e historiográficas que rodearon la publicación del libro. Después, se presenta el contenido del trabajo en relación con los siguientes estudios históricos en historia de la Reconstrucción e historia educativa. Finalmente, intenta situar el libro en una corriente más amplia de trabajo en historia de los afroamericanos.

Fuente: Urban, Wayne J. «Horace Mann Bond's Negro education in Alabama», *History of Education Quarterly*, vol 27, No. 3, Fall 1987. Reimpreso con permiso de Wayne J. Urban.

vida. El archivo de documentación de Bond proporcionó los documentos para examinar brevemente «las circunstancias personales e historiográficas que rodearon la publicación de los libros» (Urban, 1987, p. 365). La lógica del planteamiento del objetivo implica cinco preguntas de investigación que se centran en las particularidades personales de Bond, en sus circunstancias historiográficas, historia de la Reconstrucción, historia educativa e historia de los afroamericanos.

LA IMPORTANCIA DE LA SELECCIÓN DEL PROBLEMA

La importancia o **significación del problema** es la base fundamental para un estudio. Le dice al lector por qué el estudio es importante e indica las razones de la elección por parte de los investigadores de un problema determinado. Debido a que la investigación requiere conocimiento, técnicas, planificación, tiempo y recursos económicos, el problema a investigar debe tener importancia. En otras palabras, el estudio debería suponer un beneficio potencial.

Un problema de investigación es significativo cuando tiene relación con una teoría, conocimiento o práctica educativa en desarrollo, y se incrementa cuando se dan distintas causas para la investigación. La importancia puede basarse en uno o más de los siguientes criterios: si proporciona conocimiento sobre una práctica permanente, pone a prueba una teoría, es generalizable, aumenta nuestra comprensión de un fenómeno más amplio, ayuda al avance metodológico, está relacionado con un tema actual, evalúa una práctica específica en un lugar determinado o se trata de un estudio exploratorio.

1. **Conocimiento acerca de una práctica permanente.** El estudio puede proporcionar conocimiento sobre una práctica educativa permanente. Puede que se haya realizado investigación previa sobre esta práctica, pero este problema concreto de investigación no ha sido estudiado con anterioridad. La práctica estudiada puede ser común a muchos colegios pero no tiene porque encontrarse necesariamente en todos ellos. El estudio añade conocimiento sobre una práctica permanente habitual.

2. **Comprobación de una teoría.** El estudio puede poner a prueba una teoría existente con un diseño de verificación. El centro de interés puede estar constituido por las teorías de ciencias sociales acerca del desarrollo infantil o adulto, el desarrollo de la organización, los conflictos o asuntos similares. Las teorías educativas pueden estar centradas en el currículum, los modelos de instrucción, los estilos de aprendizaje, el desarrollo docente, las estrategias de enseñanza o asuntos similares. Al poner a prueba una teoría en situaciones diferentes o sobre poblaciones diferentes, el investigador puede modificarla o verificarla.

3. **Generalizabilidad.** El estudio puede ser diseñado para que los resultados puedan ser generalizables a diferentes poblaciones o prácticas. Un estudio puede repetir, o incluir, otras variables no examinadas en la investigación previa. El diseño propuesto puede sugerir el uso de una población diferente a la de la investigación original para aumentar su generalizabilidad.

4. **Ampliación de la comprensión.** Muchos estudios cualitativos realizados en el marco de la tradición fenomenológica proporcionan una ampliación de la comprensión antes que generalizabilidad. Mediante la descripción detallada de un caso de una situación social, estos estudios facilitan la comprensión de los fenómenos observados. La comprensión de los fenómenos proporciona una imagen o una configuración de expectativas razonables que podrían resultar útiles en situaciones similares.

5. **Avance metodológico.** El estudio puede incrementar la validez y fiabilidad de una prueba o emplear una metodología diferente a la de los estudios previos. Gran parte de la investigación psicométrica en educación estudia cuestiones relacionadas con la evaluación, como por ejemplo, procedimientos de administración de tests, el orden de los ítems en un documento, el formato de un artículo o la tendencia de respuesta o los procesos de información de un encuestado. Otros estudios pueden desarrollar una técnica estadística o metodológica y establecer su utilidad para la investigación.

6. **Temas actuales.** El estudio puede centrarse en un tema social o de interés inmediato. Como mencionamos previamente, movimientos políticos organizados tales como el de defensa de los derechos de las mujeres y de los derechos civiles han generado investigación educativa. El reconocimiento público de los problemas sociales ha llevado, frecuentemente, a una evaluación de sus efectos educativos. Por ejemplo, el creciente fenómeno de las familias monoparentales ha

planteado preguntas sobre el impacto de este tipo de familia en la autoestima y el rendimiento de los alumnos. Se han elaborado estudios sobre los efectos de la excesiva exposición a la televisión de los estudiantes y el empleo de las calculadoras y ordenadores en los programas educativos debido a la preocupación social ante una sociedad altamente tecnificada.

7. Evaluación de una práctica o política específicas en un lugar determinado. El estudio puede valorar una práctica o política específicas para la toma de decisiones en un lugar determinado o para grupos externos. Como apuntamos en el capítulo 1, la investigación evaluativa determina el mérito: ¿funciona la práctica y cómo puede ser mejorada en la situación concreta? También determina el valor de la práctica: ¿la práctica es efectiva?, ¿merece la pena lo que cuesta?, ¿queremos extender su uso?, ¿satisface nuestras necesidades? En los estudios políticos se tratan cuestiones similares. Este tipo de investigación suministra información para su utilización inmediata en el lugar mismo de la toma de decisiones, lo que puede ocurrir en el ámbito local, estatal o nacional. Aunque el estudio no se ocupe de la generalizabilidad o del desarrollo de la teoría, puede tener implicaciones para el desarrollo de dichos conocimientos.

8. Investigación exploratoria. La investigación exploratoria se realiza habitualmente en nuevas áreas de estudio. Pueden ser estudios cuantitativos o cualitativos. Por ejemplo, un estudio de campo podría poner a prueba un formato de examen particular para determinar su posible aplicación a los alumnos de sexto curso y comprobar si discrimina a un grupo de alumnos determinado. Los estudios exploratorios cualitativos examinan, a menudo, fenómenos que no han sido estudiados con anterioridad. La mayoría de los extractos de investigación cualitativa en este capítulo son ejemplos de investigación exploratoria. Algunos estudios exploratorios desarrollan teoría.

NORMAS DE ADECUACIÓN DE LOS ENUNCIADOS DEL PROBLEMA

Los problemas de investigación pueden ser evaluados por medio de los tres elementos descritos en este capítulo: el enunciado del problema general de investigación, la importancia del problema y el objetivo, preguntas o hipótesis específicas de la investigación. Además se pueden aplicar otros criterios.

Problema general de investigación

Las siguientes cuestiones evalúan el planteamiento general del problema:

1. ¿Implica el enunciado del problema general de investigación la posibilidad de investigación empírica?
2. ¿Restringe el planteamiento del problema el ámbito del estudio?

3. ¿Ofrece la formulación del problema el contexto educativo en el que reside el problema?

Importancia del problema

Los lectores evalúan la importancia o significación del problema en relación con uno o varios de los siguientes criterios:

- desarrolla conocimiento de una práctica habitual
- desarrolla teoría
- generalizable, esto es, amplía conocimiento o teoría
- proporciona una ampliación de la comprensión del problema
- ayuda al avance metodológico
- está relacionado con un tema político o social actual
- evalúa una práctica o política específica en un lugar determinado
- es investigación exploratoria

Preguntas o hipótesis de investigación específicas

En la evaluación de las preguntas de investigación cuantitativa y cualitativa se aplican diferentes criterios.

Cuantitativa

1. ¿Enuncia de forma concisa el objetivo, la pregunta o las hipótesis específicas de investigación que serán estudiados?
2. ¿Indica el nivel de especificidad que la pregunta o hipótesis es investigable? ¿Parecen las variables accesibles a las definiciones operacionales?
3. ¿Tiene precisión la lógica deductiva de una pregunta o hipótesis de investigación? ¿Están identificadas las variables dependientes e independientes?
4. ¿Indica la pregunta o hipótesis de investigación el marco para presentar los resultados?

Cualitativa

1. ¿Exponen las preguntas de investigación, los problemas pronosticados o el planteamiento resumido del problema el caso particular de los fenómenos a estudiar?
2. ¿Es apropiada la metodología cualitativa para la descripción de hechos presentes o pasados?
3. ¿Es razonablemente explícita la lógica inductiva de la investigación?
4. ¿Expone el propósito de la investigación el marco para presentar los resultados?

Otros criterios para normas de adecuación

Antes de llevar a cabo un estudio, el investigador, una posible institución patrocinadora, los comités de revisión y otros grupos también evalúan el problema de acuerdo con otros criterios adicionales. Estos criterios tienen que ver con la

habilidad del investigador para realizar el estudio y la viabilidad y aspectos éticos del diseño de investigación. Las preguntas características son éstas:

1. ¿Tiene el investigador un interés vital en el problema y es un tema del cual el investigador posee tanto conocimiento como experiencia?
2. ¿Son factibles el problema y el diseño en términos de medición, acceso al caso, muestra o población, permiso para manejar datos documentados, margen de tiempo para su finalización, recursos financieros y asuntos similares?
3. ¿Posee el investigador suficiente habilidad para llevar a cabo la investigación propuesta y para analizar e interpretar los resultados?
4. ¿Asegura la investigación propuesta la protección de daños o malestar mental o físico a los individuos? ¿Se proporciona el derecho al informe de consentimiento? ¿Se seguirán las prácticas éticas de investigación?

RESUMEN

Las siguientes frases resumen los aspectos principales de los enunciados de un problema de investigación, su formulación en investigación cuantitativa y cualitativa, la significación y las normas de adecuación para el enunciado de un problema.

1. Un problema de investigación implica la posibilidad de una investigación empírica.
2. Las fuentes para los problemas de investigación son las observaciones casuales, la teoría, la bibliografía relacionada, los temas actuales, las situaciones prácticas y las perspectivas personales.
3. El enunciado de un problema de investigación especifica el asunto central, el contexto educativo, la importancia y el marco para presentar los resultados.
4. En investigación cuantitativa se emplea la lógica deductiva para seleccionar el constructo, las variables y las definiciones operacionales.
5. Un constructo es una abstracción compleja que no es directamente observable.
6. Una variable es un suceso, categoría, conducta o atributo que expresa un constructo y que, dependiendo de cómo se use en un estudio concreto, adquiere valores diferentes.
7. Las variables pueden ser categóricas o continuas. Una variable puede ser dependiente, independiente, manipulada, experimental, predictora o de criterio dependiendo del diseño.
8. Una definición operacional asigna significado a una variable especificando las actividades u operaciones necesarias para su medición, clasificación o manipulación.
9. Para formular un problema cuantitativo, un investigador escoge las variables, la población y la lógica deductiva del diseño.

10. Las preguntas específicas de investigación cuantitativa pueden ser preguntas descriptivas, de relación, de diferencia o hipótesis.
11. Una hipótesis de investigación debería definir la relación o diferencia esperada entre dos o más variables, ser comprobable y ofrecer una explicación basada en una teoría o investigación previa.
12. En investigación cualitativa el tema general, el caso y la metodología están interrelacionados y seleccionados interactivamente en lugar de en etapas separadas de investigación.
13. Un caso es una situación particular seleccionada por el investigador en la que algún fenómeno será descrito a través de los significados de los participantes en hechos y en procesos.
14. Un estudio cualitativo emplea la lógica inductiva para usar registros de campo con el objeto de dar forma a una narración descriptiva y desarrollar abstracciones a partir de esa narración.
15. Los registros de campo cualitativos, obtenidos a lo largo del tiempo, se registran como notas de observación participante, transcripciones de entrevistas en profundidad o notas del investigador de documentos históricos.
16. Las descripciones cualitativas son narraciones detalladas de personas, sucesos y procesos que enfatizan los significados para los participantes.
17. Los problemas de investigación cualitativa son reformulados varias veces durante la recogida de datos, mientras que los problemas de investigación cuantitativa están definidos antes de la recogida de datos.
18. Un problema de investigación es significativo si proporciona conocimiento sobre una práctica permanente, pone a prueba una teoría, incrementa la generalizabilidad o amplía los conocimientos empíricos, ayuda al avance metodológico, se centra en un tema actual, evalúa una práctica o política específicas o se trata de un estudio exploratorio.
19. Los enunciados del problema son juzgados por criterios de formulación de un problema de investigación, su importancia, las preguntas o hipótesis específicas de investigación y la lógica y viabilidad apropiadas.

EJERCICIOS DE AUTOAPRENDIZAJE

Preguntas

1. ¿Cuál de los siguientes enunciados está redactado como un problema de investigación? El propósito de este estudio es determinar...
 - a. si se debe cambiar la política de ascensos.
 - b. la verdad de la afirmación de que la educación norteamericana alienta un sistema de clases sociales en los Estados Unidos.

- c. cómo pueden superar los estudiantes la ansiedad frente a un examen.
 - d. si hay una diferencia en la ganancia de la puntuación de rendimiento lector entre estudiantes comparables que aprendieron habilidades de análisis de palabras y aquéllos que aprendieron habilidades de comprensión.
2. El enunciado del problema de investigación proporciona...
 - a. el contexto educativo del estudio.
 - b. el marco para presentar los resultados.
 - c. la importancia del estudio.
 - d. todas las anteriores son correctas.
 3. La formulación de un problema cuantitativo requiere...
 - a. el uso de lógica deductiva para el problema.
 - b. la selección de un constructo, variables y definiciones operacionales.
 - c. la selección de una población y/o muestra.
 - d. todas las anteriores son correctas.
 4. Las preguntas de investigación cuantitativa pueden ser redactadas para indicar...
 - a. un estudio descriptivo de la situación actual de un grupo.
 - b. un estudio de relación que pronostique la influencia de una variable sobre otra.
 - c. un estudio comparativo entre dos o más conjuntos de datos.
 - d. todas las anteriores son correctas.
 5. ¿Cuál de los siguientes enunciados es incorrecto con respecto a la hipótesis de investigación? Una hipótesis de investigación...
 - a. es confirmada o no confirmada.
 - b. está relacionada con variables que pueden ser medidas, manipuladas o clasificadas.
 - c. es más específica que el planteamiento del problema.
 - d. es igual que una hipótesis estadística.
 6. ¿Cuál de las siguientes no es una definición operacional?
 - a. Las evaluaciones de dibujos de tres profesores de arte.
 - b. Los tests de «habilidades básicas» de IOWA.
 - c. Las habilidades de solución de problemas de pensamiento divergente.
 - d. La asignación de los alumnos a un curso de investigación recurriendo al criterio de haber recibido o no un curso de estadística.
 7. De los criterios para enunciar correctamente una hipótesis de investigación, cuál es el más incumplido en la hipótesis siguiente:
Los alumnos de un programa educativo vocacional contribuirán más con la sociedad que aquéllos que no están matriculados en el programa.
 - a. Una hipótesis es concisa.
 - b. Merece la pena poner a prueba la hipótesis.

- c. Una hipótesis puede ser definida operacionalmente.
- d. Una hipótesis es lógicamente precisa.
- 8. La formulación de un problema cualitativo requiere...
 - a. la selección de un caso y una metodología particular.
 - b. el empleo de lógica inductiva para construir una narración descriptiva a partir de los registros de campo cualitativos.
 - c. un acceso físico y/o psicológico a situaciones sociales del pasado o del presente.
 - d. todas las anteriores son correctas.
- 9. Un caso en investigación cualitativa se refiere a...
 - a. una persona para un estudio en profundidad.
 - b. una colección de archivos o una biblioteca jurídica.
 - c. una situación particular seleccionada por el investigador a través de la que se podría adquirir una comprensión de algunos fenómenos más amplios.
 - d. una variable.
- 10. Los problemas cualitativos...
 - a. se reformulan varias veces después de que haya empezado la recogida de datos.
 - b. normalmente, se definen como problemas generales pronosticados que se resumirán más tarde.
 - c. se centran en el «qué», «quién», «cómo» y «porqué» de las situaciones sociales.
 - d. todas las anteriores son correctas.
- 11. ¿Cuál de las siguientes frases no resulta interesante como problema de investigación?
 - a. La evaluación de una práctica o política específicas en un lugar determinado.
 - b. La ampliación o generalizabilidad de la investigación previa o extensión de los conocimientos en investigación exploratoria.
 - c. Un enfoque que no está relacionado con la investigación previa o el fenómeno general.
 - d. Un tema actual.

Problemas aplicados

Conteste a las preguntas para cada problema.

1. A continuación se incluyen unos ejemplos de temas de investigación. Indique las decisiones necesarias para realizar el estudio y repita cada una como una pregunta útil de investigación.
 - a. Efectos de formas diferentes de aprender estudios sociales.

- b. Efectos de la enseñanza cooperativa frente a la competitiva en actitudes hacia el aprendizaje.
 - c. Opiniones de los padres hacia la educación.
 - d. Características familiares y asistencia a la escuela.
 - e. Validez del WISC para el rendimiento escolar.
 2. Escriba una hipótesis direccional para el siguiente enunciado de problema e identifique el tipo de variables en la hipótesis. «Los alumnos de bajo rendimiento frecuentemente responden positivamente a los programas de modificación de conducta. ¿Existe alguna relación entre el tipo de recompensa (tangible o intangible) y el volumen de aprendizaje?»
 3. Defina una hipótesis basada en cada una de las siguientes preguntas de investigación enumeradas a continuación:
 - a. ¿Cuál es el efecto de los estudios sociales estructurados e individualizados sobre los alumnos de secundaria?
 - b. ¿Hay alguna diferencia en la implicación de los alumnos en las tareas cuando un profesor las emprende con una motivación positiva o cuando adopta una actitud indiferente ante ellas?
 - c. ¿Mejora su ajuste social la no promoción de los alumnos de primaria?
 - d. ¿Son diferentes las percepciones de estrés laboral de los maestros entre los que trabajan con niños con discapacidad, con retrasos medios o moderados?
 4. Enuncie las variables dependientes e independientes más probables para cada investigación presentada a continuación o defina que no hay variables identificables.
 - a. «La calificación liberal mejora la evaluación del profesorado».
 - b. «Comportamiento en el aula de alumnos de primaria: percepciones de los directores».
 - c. «Relación de los estilos de enseñanza del profesor con el rendimiento académico de los alumnos».
 - d. «Ideas de los movimientos de educación sexual temprana en Estados Unidos, 1890-1920».
 - e. «Reducción de la interferencia entre frases a través de ayudas contextuales».
 - f. «Efecto de dos programas de intervención basados en la escuela en los síntomas depresivos de preadolescentes».
 5. ¿Cuáles de los títulos de las investigaciones enumerados anteriormente define también claramente a los sujetos?
 6. En los siguientes enunciados de problema cualitativo, identifique el caso a estudiar.

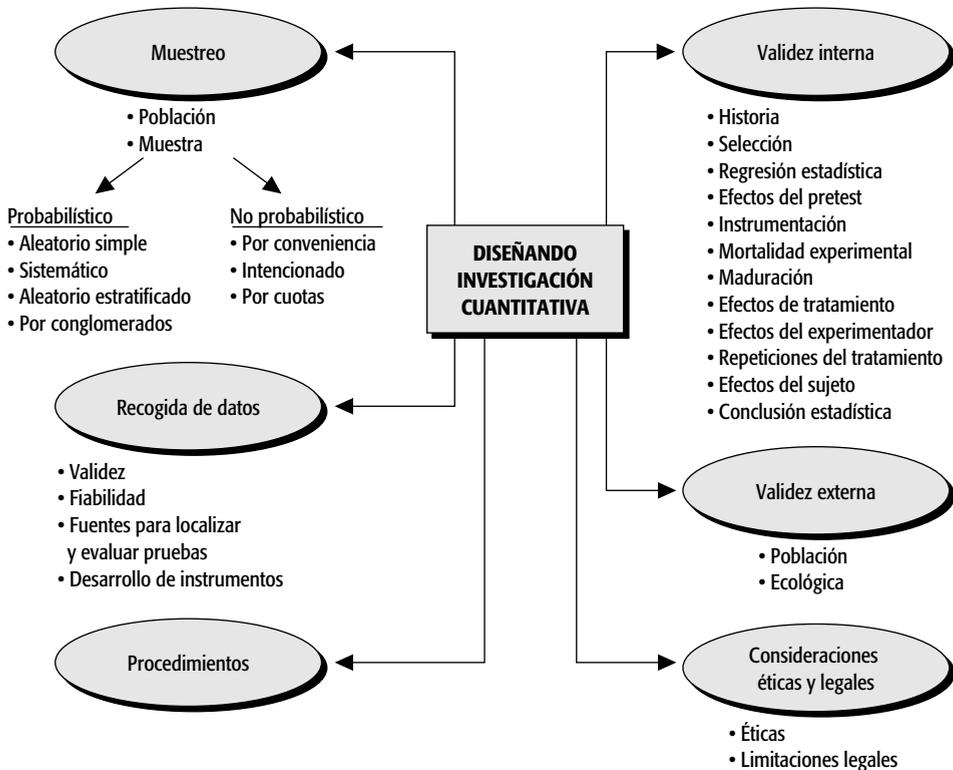
- a. Este estudio describe y analiza cómo los miembros femeninos del claustro de una universidad urbana perciben e integran sus vidas profesionales y personales.
- b. Se analizaron los registros de la junta de dirección de una escuela de un sistema escolar suburbano utilizando las ideologías articuladas por varios miembros de la junta para legitimar las políticas desde 1950 hasta 1980.
- c. El problema de investigación es describir cómo la Sra. Sue, durante su primer año como maestra de primaria, aprende su papel profesional con los alumnos, con el profesorado, con los jefes de estudios y directores y los padres, y cómo crea su «significado» acerca de la profesionalidad del profesor.
- d. El problema de investigación es describir y analizar un sistema social en la implantación de un programa de innovación en una escuela secundaria para proponer una teoría fundamentada.



Diseños y métodos de investigación cuantitativa

La parte II presenta los diseños y métodos de investigación cuantitativa. Empezamos presentando los principios fundamentales de muestreo, medida y diseño de investigación. A continuación, repasamos los conceptos y procedimientos estadísticos descriptivos que son esenciales para entender los estudios cuantitativos. También repasamos con algún detalle la validez y la fiabilidad de las puntuaciones, las fuentes a usar para evaluar las pruebas y las formas diferentes de recogida de datos cuantitativos. Estos principios resultan útiles para contestar preguntas como, por ejemplo: ¿cómo pueden resumir los investigadores grandes cantidades de datos? ¿Por qué es importante la variabilidad de los resultados? ¿Cómo pueden medirse las relaciones? ¿Qué tipo de evidencias usan los investigadores para hacer deducciones válidas a partir de las respuestas de los sujetos? ¿Cómo establecen la fiabilidad los investigadores? ¿Cuáles son las ventajas y desventajas de los cuestionarios comparados con las entrevistas? ¿Cómo se diseña un cuestionario? ¿Cómo realiza una investigación observacional o una investigación mediante encuesta? Los dos siguientes capítulos resumen las bases de los diseños de investigación cuantitativa. El capítulo 5 trata sobre los diseños descriptivo, comparativo, correlacional, predictivo y *ex post facto*, mientras que el capítulo 6 abarca los diseños experimental y de caso único. El último capítulo de esta parte es una introducción conceptual a la estadística inferencial. La intención es ayudar a la comprensión de la lógica de la probabilidad como aplicación para poner a prueba hipótesis y preguntas de investigación. Presentamos la terminología y los procedimientos estadísticos básicos para permitir al lector interpretar la sección de resultados de una investigación cuantitativa y al investigador seleccionar los procedimientos estadísticos apropiados, basados en el diseño del estudio.

Introducción al diseño de investigación cuantitativa



PALABRAS CLAVE

diseño de investigación
credibilidad
variabilidad
validez interna
validez externa
hipótesis rivales posibles
sujetos
muestra
población
muestreo probabilístico
muestreo aleatorio
muestreo aleatorio simple
muestreo sistemático
muestreo aleatorio estratificado
muestreo proporcional
muestreo no proporcionado
muestreo por conglomerados
muestreo no probabilístico
muestreo por conveniencia
muestreo intencional
muestreo por cuotas
validez del test
fiabilidad del test
historia
selección

regresión estadística
pretest
instrumentación
mortalidad experimental
maduración
difusión del tratamiento
efectos del experimentador
repeticiones del estudio
efectos del sujeto
características de la demanda
validez externa de población
validez externa ecológica
efecto Hawthorne
validez de constructo
informe de consentimiento

Diseñar investigación cuantitativa supone elegir sujetos, técnicas de recogida de datos (cuestionarios, observaciones o entrevistas), procedimientos para la recogida de datos y la implantación de tratamientos. Todos estos elementos constituyen la parte metodológica del estudio. En este capítulo se presentarán los elementos esenciales del diseño de investigación cuantitativa, destacando los principios más importantes para la elaboración de conceptos y la planificación de un estudio. Discutiremos cada uno de estos componentes, prestando especial atención a los principios que mejoran la calidad de la investigación. También trataremos las consideraciones éticas y legales que son importantes en la planificación y realización de una investigación.

EL PROPÓSITO DEL DISEÑO DE INVESTIGACIÓN

Un **diseño de investigación**¹ se refiere a un plan para la selección de sujetos, de escenarios de investigación y de procedimientos de recogida de datos que respondan a las preguntas de la investigación. El diseño expone qué individuos serán estudiados, cuándo, dónde y bajo qué circunstancias. La meta de un diseño de investigación sólido es proporcionar resultados que puedan ser considerados creíbles. La **credibilidad** se refiere al grado en el que los resultados se aproximan a la realidad y pueden ser considerados fiables y razonables. La credibilidad aumenta cuando el diseño de investigación tiene en cuenta las posibles fuentes de error que pueden debilitar la calidad de la investigación y distorsionar los resultados. El objetivo de un buen diseño de investigación es, por lo tanto, proporcionar una respuesta fiable a una pregunta, y el error reduce la credibilidad de los resultados. Cuando se diseña cuidadosamente el estudio, el investigador puede eliminar o, al menos, reducir las fuentes de error. Todas las fuentes potenciales de error en una investigación no pueden ser controladas por completo, pero hay principios en la planificación de investigación para minimizar tales influencias.

Fuentes de variabilidad

Uno de los principios importantes de la investigación cuantitativa es considerar las diferentes *fuentes de variabilidad*. La **variabilidad** hace referencia a cuántas observaciones sobre la misma cuestión adquieren valores diferentes. Por ejemplo, sabemos que nuestro estado de ánimo varía de un día para otro, del mismo modo que conocemos que el rendimiento académico de un alumno no será el mismo cada vez que realice un examen.

Desde el punto de vista del diseño, es importante reconocer y controlar tres fuentes de **variabilidad**: sistemática, error y externa. La variabilidad sistemática está relacionada con las variables que están siendo investigadas. Lo que el investigador busca es un diseño que maximice este tipo de variación. Por ejemplo, si estudiamos la relación entre tiempo y rendimiento, diseñaríamos la investigación de tal manera que las dos variables de interés, implicación y rendimiento, posean una variabilidad alta. Si, por ejemplo, todos los estudiantes obtienen puntuaciones de rendimiento iguales o muy parecidas, no seríamos entonces capaces de demostrar la relación.

¹ El término diseño de investigación se utiliza, a menudo, para referirse a tipos específicos de diseños denominados diseños «preexperimentales» y «semiexperimentales» por Campbell y Stanley (1963). La importancia de este libro está en describir los fundamentos del control del error en todos los aspectos de la recogida de información, que incluye pero no está limitado a los diseños de Campbell y Stanley. Los principios descritos son valiosos para la investigación no experimental así como para los estudios que infieren relaciones causales

De forma similar, en un experimento, se busca maximizar la variación de la variable dependiente cuando se comparan los grupos. Esto se consigue, a menudo, asegurándose de que los tratamientos en el estudio producirán potencialmente resultados bastante diferentes. Por ejemplo, es probable que la varianza sistemática sea mayor en un estudio que compara la enseñanza individualizada con un pequeño grupo de estudio que si compara dos tipos de enseñanza de pequeños grupos de discusión.

La varianza de error debe ser minimizada. Incluye el error de muestreo, de medida y otros tipos de hechos aleatorios que hacen difícil demostrar relaciones. La varianza externa necesita ser controlada. Este tipo de variabilidad afecta a las relaciones de modo directo antes que aleatorio. Por ejemplo, en el estudio de la relación entre las puntuaciones de un test y el tamaño de la clase, el nivel socioeconómico de los estudiantes sería una variable que debería controlarse. Esto es, se obtendría una estimación más ajustada de la relación si se controlaran, estadísticamente, las repercusiones del nivel socioeconómico, qué están relacionadas con el rendimiento.

Validez del diseño

En el contexto del diseño de investigación, el término **validez** significa el grado en que las explicaciones científicas de sucesos coinciden con la realidad. La validez hace referencia a la verdad o falsedad de las proposiciones generadas por la investigación. Las explicaciones sobre los sucesos observados se aproximan a la realidad o verdad, y el grado en el que las explicaciones son precisas determina la validez de la investigación. Se reconocen dos formas de validez de diseño en investigación cuantitativa. La **validez interna** expresa el grado en que las variables externas se han tenido en cuenta o se han controlado. La **validez externa** se refiere a la generalizabilidad de los resultados, el grado en que los resultados y las conclusiones pueden ser generalizados a otras personas y ambientes. En algunos estudios, como en la mayoría de la investigación aplicada, hay una intención clara de generalización a otras personas y ambientes, mientras que en otros estudios no se concibe o no es posible la generalización más allá de las personas, el tiempo y el contexto de la investigación. Tanto la validez interna como la externa son conceptos importantes que deben tenerse muy claros al diseñar una investigación, por lo que examinaremos minuciosamente cada uno de ellos.

Al diseñar o leer investigación cuantitativa es necesario considerar quién será evaluado (sujetos), por medio de qué serán evaluados (instrumentos/ pruebas) y cómo serán evaluados (procedimiento para la recogida de datos); y para diseños experimentales, cómo serán administrados los tratamientos experimentales. Por lo tanto es importante preguntar: ¿ha ocurrido o se ha hecho algo que pudiera proporcionar una explicación de los resultados por medio de una hipótesis

alternativa? El término «alternativo» se emplea en el sentido de que se puede añadir a la hipótesis enunciada o del objetivo de la investigación. (Por ejemplo, una hipótesis alternativa al estudio de si fumar causa cáncer de pulmón es que la alimentación puede contribuir al cáncer de pulmón). Esta cuestión representa la búsqueda de variabilidad externa. Campbell y Stanley (1963) hacen referencia a tales explicaciones como **hipótesis alternativas posibles**. La búsqueda de hipótesis alternativas posibles es esencial para asegurar la calidad de la investigación. Consideremos, por ejemplo, las siguientes preguntas. Cada una se aplica a una posible fuente de error que puede conducir a una hipótesis alternativa que podría explicar los resultados.

1. ¿Tiene el investigador algún sesgo con respecto a los sujetos o al asunto investigado?
2. ¿Son conscientes los sujetos de que están siendo estudiados?
3. ¿Están respondiendo honestamente los sujetos?
4. ¿Han recibido ambos grupos los tratamientos como se describe?
5. ¿El sexo del entrevistador ha marcado una diferencia?
6. ¿Abandonaron muchos sujetos antes del final del estudio?
7. ¿Afectó la hora del día de la investigación a los resultados?

Se utilizan varias técnicas para reducir la varianza externa en investigación cuantitativa: elección al azar de sujetos, mantenimiento de condiciones o factores constantes, construcción de condiciones o factores dentro del diseño como variables independientes y ajustes estadísticos (Wiersma, 2000). La elección al azar es deseable tanto en la selección de sujetos a partir de poblaciones numerosas como en la asignación de sujetos a grupos para investigar el efecto de una variable en otra. La selección aleatoria, que será descrita en detalle más adelante en la siguiente sección, permite al investigador generalizar los resultados más allá del grupo inmediato estudiado. La asignación aleatoria ayuda al control del error asociado a características de sujetos en diferentes grupos.

Si el investigador considera que las condiciones de la recogida de datos podrían afectar a los resultados, el estudio puede ser diseñado para asegurar que todas las condiciones sean lo más semejantes que sea posible. Por ejemplo, en un estudio de observación sobre la relación entre la conducta del profesor y la atención del alumno, tanto la hora del día en que el observador registra los datos como la asignatura (mañanas frente a tardes, o matemáticas frente a historia) podría marcar una diferencia en la atención del alumno. Una forma de controlar esta fuente potencial de error es asegurarse de que todas las observaciones se hacen a la misma hora, durante clases de la misma asignatura. En este ejemplo, el investigador, también, podría controlar estas influencias potenciales convirtiéndolas en variables independientes. Esto podría lograrse asignando observadores a cada asignatura de interés y examinando cada una de éstas tanto por la mañana como

por la tarde. De este modo el investigador puede evaluar el efecto de la hora del día y de la asignatura en lugar de controlarlo simplemente.

En los estudios cuantitativos, el control de las posibles variables externas es esencial, aunque la investigación educativa raramente exhibe el grado de control que se puede constatar en los estudios de fenómenos físicos o en psicología. De este modo, el investigador debe buscar constantemente factores (variables externas) que puedan tener influencia en los resultados o conclusiones del estudio. Para la investigación cuantitativa el concepto de validez interna describe la eficacia con la que se han controlado las variables externas. La preocupación reside en el modo en que el procedimiento, el muestreo de sujetos y las pruebas afectan al grado en que las variables externas están presentes para complicar la interpretación de los resultados. Un estudio potente o sólido en validez interna controla con éxito todas o la mayoría de las variables externas de tal manera que el investigador puede estar seguro de que, por ejemplo, X causó cambios en Y. Estudios poco sólidos en validez interna son difíciles de interpretar ya que es imposible discernir si los resultados se deben a la variable independiente o a alguna variable externa que no fue controlada o tenida en cuenta. Es importante para los investigadores ser conscientes de los factores habituales que pueden convertirse en variables externas y elaborar conceptos y leer la investigación pensando en ellos. Puesto que el control completo de las variables externas en investigación educativa resulta difícil, por no decir imposible, todas las amenazas a la validez interna que no puedan ser atajadas deberían ser tomadas en cuenta para la interpretación de los resultados.

SUJETOS: POBLACIONES Y MUESTRAS

Uno de los primeros pasos en el diseño de investigación cuantitativa es elegir los sujetos. Los **sujetos** (abreviado como S) son los individuos que participan en el estudio; es decir, a partir de los que se recogen los datos. Por ejemplo, en un experimento, un sujeto es cada persona a la que se administra un tratamiento y de la que se mide una respuesta. En los estudios no experimentales, se consideran sujetos los individuos cuya conducta presente o pasada es utilizada como datos. Por ejemplo, un investigador podría manejar las notas de los alumnos de décimo curso en 1995 y cada alumno de décimo que proporcionaba sus notas sería considerado un sujeto.

A los sujetos considerados como grupo se les denomina, normalmente, **muestra**. La muestra puede seleccionarse a partir de un grupo mayor de personas, conocido como población, o puede referirse simplemente al grupo de sujetos a partir de los cuales se han recogido los datos (aunque los sujetos no estén seleccionados de una población). La naturaleza del procedimiento de muestreo que se emplea en un estudio determinado, habitualmente, se describe con uno o más adjetivos, como por ejemplo muestreo aleatorio, muestreo por conveniencia o

muestreo estratificado. Esto proporciona información sobre la técnica utilizada para formar la muestra. Consideraremos dos categorías principales de diferentes técnicas de muestreo: probabilístico y no probabilístico. Sin embargo, es necesario primero describir más el concepto de población.

¿Qué es una población?

Una **población** es un grupo de elementos o casos, ya sean individuos, objetos o acontecimientos, que se ajustan a criterios específicos y para los que pretendemos generalizar los resultados de la investigación. Este grupo también se conoce como *población objetivo* o *diana* o *universo*. La población objetivo, a menudo, es diferente de la lista de elementos a partir de la cual se selecciona realmente la muestra, que se denomina *población del estudio* o *marco de muestreo*. Por ejemplo, en un estudio sobre profesores principiantes, la población objetivo puede ser los profesores en su primer año de trabajo a lo largo de Estados Unidos, en todas las clases de escuelas. La población del estudio puede ser una lista de los profesores en su primer año de trabajo de 24 estados. De este modo, aunque lo que intenta la investigación es generalizar a todos los profesores principiantes, la estructura de la muestra establece algunas limitaciones a tales generalizaciones.

Es importante para los investigadores definir completa y cuidadosamente tanto la población objetivo como el marco de muestreo. El proceso comienza con el problema de investigación y la revisión bibliográfica, a través de los que se describe una población conceptualmente o en términos generales. Se necesita, después, una definición más específica, basada en las características demográficas como por ejemplo, edad, género, localización, nivel, puesto de trabajo y época del año. Estas características, a veces, se denominan *variables de identificación*. Por ejemplo, en un estudio sobre los alumnos que forman parte de una minoría de primer curso en un entorno rural, hay cuatro variables de identificación: rural, alumnos, primer curso y minoritario. Se incluye, luego, una descripción completa en el apartado de sujetos del informe.

Muestreo probabilístico²

En el **muestreo probabilístico**, los sujetos se extraen a partir de una población más amplia de manera que se conoce la probabilidad de selección de cada miembro de la población, aunque las probabilidades no son necesariamente iguales. Este tipo de muestreo se realiza para proporcionar estimaciones válidas de lo

² Ver R. M. Jaeger (1984), *Sampling in education and the social sciences*, Nueva York: Longman, Inc. y G. T. Henry (1990), *Practical sampling*, Newbury Park, CA: Sage, para una discusión más detallada de cómo seleccionar muestras probabilísticas.

que es cierto para una población a partir de un grupo más pequeño de sujetos (muestra). Es decir, que lo que se describe en la muestra también será cierto, con cierto grado de error, para la población. Cuando el muestreo probabilístico se realiza correctamente, puede seleccionarse sólo un porcentaje muy pequeño de la población. Esto ahorra tiempo y dinero sin sacrificar la precisión. De hecho, en la mayoría de las ciencias sociales y en Investigación Educativa no es práctico ni necesario medir todos los elementos de la población de interés.

Pueden utilizarse varios métodos de muestreo probabilístico para extraer muestras representativas o *no sesgadas* a partir de una población. Cada método requiere algún tipo de **muestreo aleatorio**, en el que cada miembro de la población total o de subgrupos de la población, poseen las mismas oportunidades que otros miembros del mismo grupo de ser seleccionados. El muestreo aleatorio evita el sesgo porque hay una alta probabilidad de que todas las características de la población estén representadas en la muestra. Sin embargo, si no se siguen los procedimientos correctos, lo que parece un muestreo aleatorio en realidad es un muestreo sesgado (sesgado en el sentido de que ciertas características de la población están supra o infrarrepresentadas). Por ejemplo, puede pensar que es posible obtener una muestra aleatoria de estudiantes universitarios plantándose en una esquina concurrida y seleccionando uno de cada tres o cuatro estudiantes. Sin embargo, puede que no haya contado de manera exacta y de forma inadvertida puede haber seleccionado más hombres que mujeres, o más estudiantes jóvenes que mayores. Tal procedimiento dará como resultado una muestra sesgada.

Dado que existen varios procedimientos de muestreo probabilístico, presentaremos los cuatro más comunes que se emplean en investigación educativa: muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y muestreo por conglomerados.

Muestreo aleatorio simple. En el **muestreo aleatorio simple**, los sujetos son seleccionados a partir de la población, de tal manera que todos los miembros de la población tienen la misma probabilidad de resultar escogidos. Este método se usa a menudo cuando la población es pequeña. Por ejemplo, un tipo común de muestreo aleatorio simple es extraer nombres de un sombrero.

Con una población mayor es necesario usar un procedimiento más preciso. Uno de esos procedimientos es utilizar una tabla de números aleatorios, que es un conjunto de números seleccionados al azar (en el Apéndice B se presenta una tabla de números aleatorios). Suponga, por ejemplo, que un investigador tenía una población de 100 alumnos de tercer curso y quería seleccionar 20 por medio del muestreo aleatorio simple. Primero, se le asigna un número de 001 a 100 a cada estudiante de la población (podría ser de 00 a 99). Segundo, el investigador selecciona al azar un punto de partida en una tabla de números aleatorios. Entonces lee los números de tres dígitos, desplazándose por las filas o por las columnas. El investigador sigue seleccionando veinte números de tres dígitos entre el 000 y el 100. La tabla 4.1

contiene un ejemplo de muestreo aleatorio simple. Están señalados con un círculo cinco de los veinte sujetos elegidos para ser incluidos en la muestra, empezando en la esquina superior izquierda y desplazándose hacia abajo por cada columna.

Una forma más eficaz, y cada vez más frecuente, de trazar una muestra aleatoria simple es a través de un programa informático adecuado. Este método es especialmente fácil y efectivo si el marco de muestreo tiene un formato electrónico.

Muestreo sistemático. En el **muestreo sistemático** se selecciona cada enésimo elemento a partir de una lista de todos los elementos de la población, comenzando con un elemento seleccionado de forma aleatoria. Suponga que es necesario confeccionar una muestra del 10% a partir de una población de 100. Se selecciona al azar un número del uno al diez como punto de partida. Si ha salido el 5, se seleccionará cada décimo número de la lista 5, 15, 25, 35, etc. Esta aproximación sólo puede usarse cuando el investigador tiene una lista secuencial de todos los sujetos de la población, pero esto es más fácil que el muestreo aleatorio simple porque no es necesario asignar un número a cada miembro de la población.

En el muestreo sistemático existe una posible limitación si la lista de casos de la población se ordena según un patrón sistemático que está relacionado con lo que se está investigando. Por ejemplo, suponga que está muestreando profesores de muchos colegios y la lista de cada colegio está ordenada por antigüedad. Si este patrón cíclico (denominado *periodicidad*) está relacionado con cada enésimo sujeto, la muestra podría excluir sistemáticamente a profesores de ciertas edades y no ser representativa de la población. Los listados alfabéticos normalmente no crean periodicidad y son más adecuados para la selección sistemática de sujetos.

TABLA 4.1: DÍGITOS ALEATORIOS VARIADOS

46614	20002	17918
16249	05217	54102
91530	62481	05374
62800	62660	20186
10089	96488	59058
47361	73443	11859
45690	71058	53634
50423	53342	71710
89292	32114	83942
23410	41943	33278
59844	81871	18710
98795	87894	00510
86085	03164	26333
37390	60137	93842
28420	10704	89412

Una ventaja del muestreo sistemático es que si la población está ordenada por una variable que está relacionada con la variable dependiente, esta ordenación tiene el efecto de estratificar y asegurarse de que la muestra está representada por cada nivel de esa variable. Por ejemplo, si la lista de población está ordenada por las puntuaciones en un test de aptitud (primero las puntuaciones más altas, seguidas por las puntuaciones más bajas), cuando se selecciona el *n*ésimo sujeto, se puede estar seguro de que todos los niveles de aptitud estarán representados en la muestra. El extracto 4.1 ilustra un muestreo sistemático. En este estudio, la primera de las tres muestras, profesores en activo, se selecciona al azar «a partir de una lista». Contraste esto con las dos siguientes muestras, que no están seleccionadas de forma aleatoria.

EXTRACTO 4.1 **MUESTREO SISTEMÁTICO Y POR CONVENIENCIA**

Se extrajeron tres muestras de este estudio. Fueron muestras de (a) profesores en activo, (b) alumnos de segundo año de carrera empezando un programa educativo de maestros y (c) alumnos de último año de carrera terminando un programa educativo de maestros (pero antes de impartir clases). Para la primera muestra, los formularios del estudio se enviaron por correo a 700 profesores en un estado rural del Oeste seleccionados de forma aleatoria a partir de la lista del Departamento de Educación Estatal de todos los profesores licenciados.

La segunda muestra fue una muestra por conveniencia de tres secciones de una clase de fundaciones educativas que habitualmente eligen los estudiantes de segundo año de la universidad que acaban de inscribirse en un programa educativo de maestros.

La tercera muestra también fue una muestra por conveniencia de cuatro secciones de una clase de psicometría que eligieron los estudiantes de último año.

Fuente: Green, K. E. (1992), «Differing opinions on testing between preservice and inservice teachers», *Journal of Educational Research*, 86, 37-42. Reimpreso con permiso.

Muestreo aleatorio estratificado. Así se denomina una variante habitual del muestreo aleatorio. En este procedimiento, la población se divide en subgrupos o estratos sobre la base de una variable elegida por el investigador, como por ejemplo, el género, la edad o el nivel educativo. Una vez que se ha dividido la población, se extraen al azar las muestras de cada subgrupo. El número de sujetos extraídos puede ser tanto *proporcional* como *no proporcional*. El **muestreo proporcional** se basa en el porcentaje de sujetos de la población que está presente en cada estrato. De este modo, si el 40% de los sujetos en la población está representado en el primer estrato, entonces, el 40% de la muestra final debería ser de ese estrato. En el **muestreo no proporcional** (o desproporcionado), el

investigador selecciona el mismo número de sujetos para cada estrato de la muestra. Si el muestreo aleatorio estratificado, ya sea proporcional o no proporcional, es a menudo más eficiente que el muestreo aleatorio simple es porque es necesario utilizar un número más pequeño de sujetos. Dividir la población en subgrupos permite al investigador comparar resultados de subgrupos.

Por ejemplo, en el extracto 4.2, los investigadores han estratificado la población de maestros sobre la base del curso y sus puntuaciones en el TFO y la población estudiantil por aula. El muestreo se representa gráficamente en la figura 4.1. Se emplea el muestreo no proporcional para asegurar que la muestra final tiene un número suficiente de sujetos en cada grupo.

EXTRACTO 4.2 **MUESTREO ALEATORIO ESTRATIFICADO**

PARTICIPANTES

Se seleccionaron de forma aleatoria treinta y seis maestras de primaria de un colectivo voluntario de un distrito escolar del sur. La muestra estaba compuesta por 18 maestras de segundo curso y 18 de quinto curso. La muestra se restringió a mujeres debido al escaso número de maestros varones de primaria en ese distrito. Basada en las puntuaciones TFO*, se seleccionaron 9 maestras de cada nivel de aquéllas que eran independientes de campo y otras 9 de las que eran dependientes de campo. Había 12 alumnos (6 niños y 6 niñas) que fueron seleccionados al azar de la clase de cada profesor para ser objeto de observación. Los alumnos de segundo tenían edades comprendidas entre los 7 años y los 7 años y 11 meses, mientras que los alumnos de quinto tenían entre 10 años y 10 años y 11 meses.

*TFO se refiere al test de las figuras ocultas (EFT, *Embedded Figures Test*)

Fuente: De Saracho, O. N. y Dayton, C. M. (1980), «Relationship of teachers' cognitive styles to pupils' academia achievement gains», *Journal of Educational Psychology*, 72, 544-549.

Muestreo por grupos. El muestreo por grupos se parece al muestreo aleatorio estratificado en el que se identifican grupos de individuos a partir de la población, y los sujetos se extraen de estos grupos. Sin embargo, en el **muestreo por grupos**, el investigador identifica, según le conviene, unidades grupales naturales, como por ejemplo vecindarios, colegios, distritos o regiones, en lugar de sujetos individuales, y selecciona al azar alguna de estas unidades para el estudio. Una vez que se han seleccionado las unidades, se escogen los individuos a partir de cada una de ellas. Por tanto, el muestreo por grupos implica dos etapas y, debido a que en la primera etapa sólo deben ser elegidos los grupos, el investigador se

ahorra el coste de la selección individual a partir de la población completa. El muestreo por grupos obtiene una muestra menos representativa de la población que el muestreo aleatorio simple o el estratificado y se utiliza muy a menudo en los casos en los que no es práctico o factible obtener una lista de todos los miembros de la población.

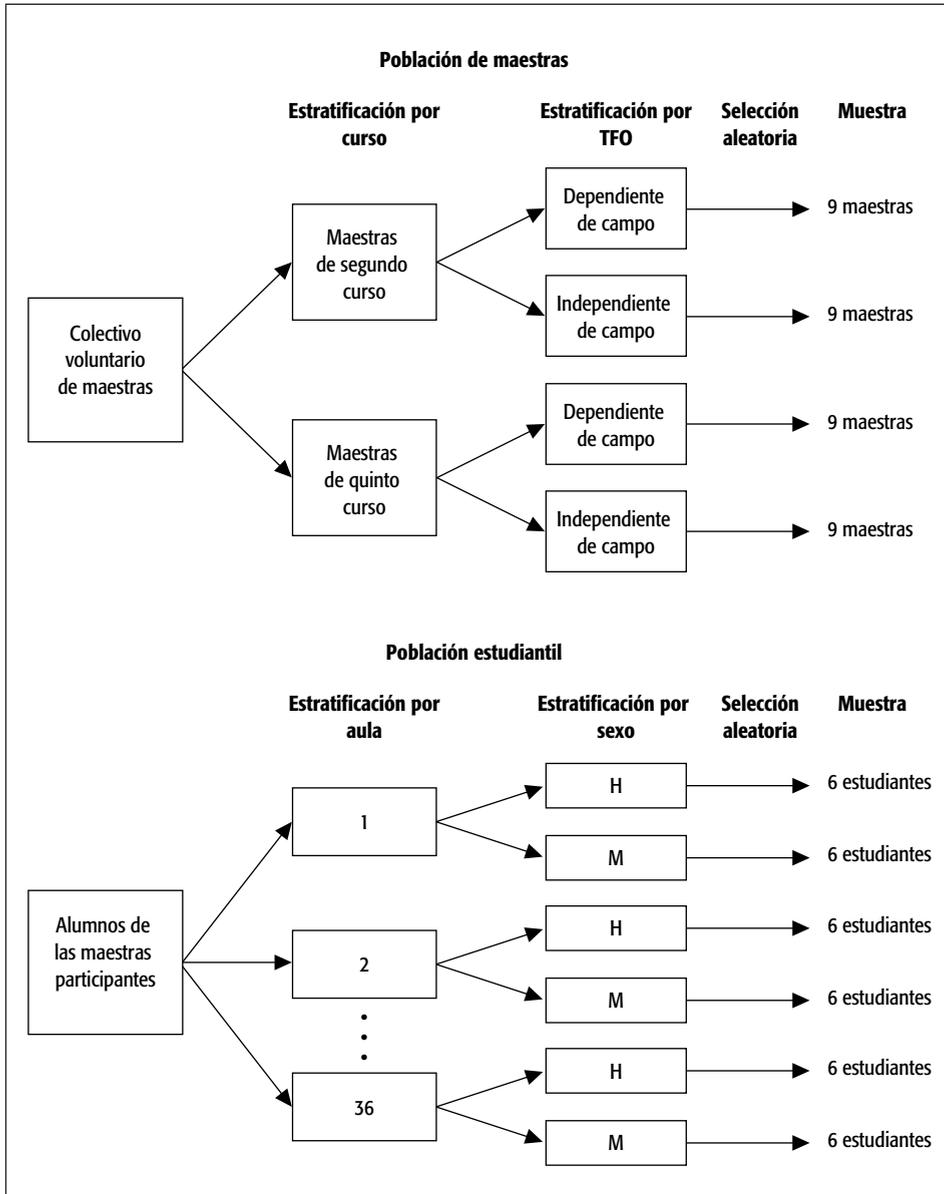
Muestreo no probabilístico

En muchos estudios educativos y, particularmente, en investigaciones experimentales y semiexperimentales, los muestreos probabilísticos no resultan apropiados, o puede ser imposible o no ser factible seleccionar sujetos a partir de un grupo mayor. En su lugar, se emplea el **muestreo no probabilístico**. De hecho, esta forma de muestreo es la más frecuente en investigación educativa. El muestreo no probabilístico no incluye ningún tipo de muestreo aleatorio. Más bien, el investigador toma sujetos que resultan accesibles o que pueden representar ciertos tipos de características. Por ejemplo, podría ser una clase de estudiantes o un grupo constituido para una reunión.

Existen muchas circunstancias que reúnen a las personas en situaciones que resultan eficaces y son aprovechadas de forma barata por la investigación. Consideraremos tres tipos de muestreo no probabilístico: muestreo por conveniencia, muestreo intencional y muestreo por cuotas. Las técnicas de muestreo no probabilístico adicionales se tratan en el capítulo 10 para diseños cualitativos.

Muestreo por conveniencia. Una **muestra por conveniencia** es un grupo de sujetos seleccionados sobre la base de ser accesibles o adecuados. Es conveniente usar los grupos como sujetos. Esto podría ser, por ejemplo, una clase universitaria de un profesor que está realizando una investigación acerca de los estilos de aprendizaje de los alumnos universitarios, aulas de profesores matriculados en una licenciatura, directores de colegio que participan en un seminario o una conferencia, gente que decide ir al centro comercial el sábado, o gente que responde a un anuncio que busca sujetos para participar en una investigación (ver extracto 4.1). Aunque este tipo de muestra facilita el llevar a cabo la investigación, tiene dos limitaciones importantes. Primero, no hay una forma precisa de generalización a partir de la muestra a cualquier tipo de población. Esto significa que la generalizabilidad de los resultados estaría restringida a las características de los sujetos. Esto no significa que los resultados no sean útiles, simplemente significa que es necesaria mucha cautela a la hora de generalizar. A menudo, los investigadores describirán muestras por conveniencia cuidadosamente para mostrar que, aunque no pudieron emplear una selección aleatoria, las características de los sujetos coinciden con las de la población o, por lo menos, con las de una parte importante de ésta.

FIGURA 4.1: SELECCIÓN ALEATORIA ESTRATIFICADA DE SUJETOS PARA EL ESTUDIO DE SARACHO Y DAYTON



Aunque es necesario que seamos muy cuidadosos con el muestreo por conveniencia, a menudo nos proporciona la única posibilidad para la investigación. Además, el propósito principal de la investigación puede ser no generalizar

sino conseguir un mejor entendimiento de las relaciones que pueden existir. En este caso, no es necesario el empleo del muestreo probabilístico. Suponga que un investigador está estudiando la relación entre creatividad e inteligencia y la única muestra posible consiste en niños de la escuela de primaria de su pueblo. Realiza el estudio y los resultados indican una relación moderada: los niños que son más inteligentes tienden a ser más creativos. Debido a que no ha habido un muestreo probabilístico, ¿deberíamos ignorar los resultados o sugerir que no son creíbles o útiles? Esta decisión parece demasiado dura. Es mucho más razonable interpretar los resultados como válidos para niños similares a los estudiados. Si la escuela se encuentra en una zona socioeconómica baja, los resultados no serán tan útiles como lo serían si hubieran estado representados todos los niveles socioeconómicos. La decisión, pues, no es descartar los resultados, sino limitarlos al tipo de sujetos de la muestra. A medida que la investigación acumula más y más muestras por conveniencia, la credibilidad global de los resultados aumenta.

Muestreo intencionado. En el **muestreo intencionado** (algunas veces denominado *muestreo intencionado* o *de juicios*) el investigador selecciona elementos particulares a partir de la población que será representativa o proporcionará información sobre el elemento de interés. Sobre la base del conocimiento que tiene el investigador de la población, se hace un juicio sobre qué sujetos deberían ser seleccionados para proporcionar la información más adecuada que responda al propósito de la investigación. Por ejemplo, en investigación sobre enseñanza eficaz, el hecho de observar profesores expertos puede proporcionar más información que una muestra de todos los profesores. Para estudiar la eficacia de la escuela puede resultar más revelador entrevistar al personal adecuado que a una muestra aleatoria de toda la plantilla.

Como veremos en el capítulo 10, existen diversos tipos de procedimientos de muestreo intencionado para investigaciones cualitativas. En los estudios cuantitativos se pone el énfasis en el juicio del investigador para seleccionar una muestra que resulte representativa de la población. Esto es, el énfasis tiende a situarse en la representatividad, mientras que los investigadores cualitativos están más interesados en la selección de casos que sean «ricos en información». El extracto 4.3 es un ejemplo del uso de un procedimiento de muestreo intencional en un estudio cuantitativo.

Muestreo por cuotas. El **muestreo por cuotas** se emplea cuando el investigador no puede utilizar el muestreo probabilístico, pero todavía es capaz de seleccionar sujetos sobre la base de las características de la población. Se definen ciertas «cuotas» para que la muestra represente a la población de acuerdo a esas características. Se identifican diferentes perfiles entre los principales grupos de la población y, luego, se seleccionan los sujetos, de forma no aleatoria, para representar a cada grupo. Por ejemplo, es frecuente establecer cuotas para

características como género, raza, edad, curso, puesto de trabajo y localización geográfica. La ventaja de este tipo de muestreo es que resulta más representativo de la población que un muestreo por conveniencia o intencional, pero todavía hay una gran dependencia del juicio del investigador para seleccionar a los sujetos.

Los muestreos no probabilísticos tienen dos limitaciones importantes. Primero, la muestra no es representativa de una población numerosa, por lo tanto la generalización es más restrictiva. La generalizabilidad de los resultados estará limitada a las características de los sujetos. Esto no quiere decir que los resultados no sean útiles, simplemente significa que se necesita cautela al realizar generalizaciones. Los investigadores describirán cuidadosamente a los sujetos para mostrar que, aunque no fueron seleccionados de forma aleatoria a partir de una población numerosa, las características de los sujetos son representativas de gran parte de la población.

EXTRACTO 4.3 **MUESTREO INTENCIONAL**

Los datos para el estudio fueron recogidos en 16 institutos de California y Michigan. Se seleccionaron intencionalmente los 16 colegios para garantizar la diversidad en los contextos de enseñanza secundaria en términos de políticas de estado, recursos de distrito, organización escolar y composición estudiantil.

Fuente: Raudenbush, S. W., Rowan, B. y Cheong y. F. (1993), «Higher order instructional goals in secondary schools: Class, teacher, and school influences», *American Educational Research Journal*, 30 523-553.

Una segunda limitación es que una muestra no probabilística puede estar sesgada. Esto es cierto particularmente para las *muestras voluntarias*, en las que los sujetos participan de forma voluntaria en la investigación. Los estudios indican que los voluntarios difieren de los no voluntarios en muchos aspectos. Rosenthal y Rosnow (1975) concluyen que, en general, los voluntarios tienden a tener niveles educativos más elevados, una clase social más alta, son más inteligentes, más sociables, menos convencionales, menos autoritarios, menos conformistas, más altruistas y más extrovertidos que los no voluntarios. Estas características podrían afectar obviamente a los resultados y llevar a conclusiones que serían diferentes si se hubiese empleado un muestreo probabilístico. Por ejemplo, suponga que un investigador ha llevado a cabo un estudio sobre las actitudes de los universitarios hacia la facultad a la que acudieron. Se enviaron cartas a 500 licenciados; 25 aceptaron regresar al campus para las

entrevistas. ¿Es razonable concluir que las actitudes de estos 25 voluntarios son representativas de la clase?

Para decidir sobre un procedimiento de muestreo, es de ayuda tener en cuenta las ventajas y desventajas de los diferentes procedimientos, descritas en la tabla 4.2. La elección final del procedimiento dependerá del propósito, la disponibilidad de sujetos y los recursos financieros del investigador.

Tamaño de la muestra

El número de sujetos en un estudio se denomina *tamaño de la muestra*, representado por la letra n . El investigador debe determinar el tamaño de la muestra que le proporcionará los datos suficientes para responder a la pregunta de investigación. La regla general para determinar el tamaño de la muestra es obtener un número suficiente para proporcionar un resultado creíble. En situaciones en las que se selecciona una muestra al azar, un tamaño muestral que constituye sólo un pequeño porcentaje de la población puede aproximarse satisfactoriamente a las características de ésta. Rowntree (1941) ilustra este punto en un estudio del porcentaje de ingresos que se gastaban en alquiler cinco categorías de familias de clase trabajadora en Inglaterra. Se recogieron datos de toda la población y se compararon con los datos que habían sido obtenidos con diferentes tamaños de muestras aleatorias. Como se indica en la tabla 4.3, existe una diferencia pequeña entre un tamaño muestral del 2% (uno de cincuenta) y 10% (uno de diez).

La determinación del tamaño de la muestra, debería tomar en consideración varios factores —el tipo de investigación, la hipótesis de investigación, las limitaciones financieras, la importancia de los resultados, el número de variables estudiadas, los métodos de recogida de datos y el grado de exactitud necesario—. El impacto de estos factores se resume más adelante³.

1. **El tipo de investigación.** La investigación correlacional debería realizarse con un mínimo de treinta sujetos, y en la investigación que compara grupos debería haber como mínimo quince sujetos en cada grupo (algunos experimentos altamente controlados poseen de ocho a diez sujetos por cada grupo). En la investigación mediante encuesta debería haber aproximadamente cien sujetos por cada subgrupo principal que se analice y de veinte a cincuenta en los grupos menores.

³ Para una discusión de los procedimientos más sistemáticos para estimar el número de sujetos necesarios dependiendo del diseño de investigación del estudio, ver H. C. Kraemer y Thielman (1987), *¿How many subjects? Statistical power analysis in research*. Newbury Park, CA: Sage y Jaeger (1984).

TABLA 4.2: VENTAJAS Y LIMITACIONES DE LOS MÉTODOS DE MUESTREO

Método de muestreo	Ventaja	Limitaciones
Probabilístico		
Aleatorio simple	Fácil de entender. Se necesita poco conocimiento sobre la población. Libre de error en la clasificación de sujetos. Facilidad para analizar e interpretar los resultados.	Requiere numerar cada elemento de la población. Mayor error de muestreo que en el muestreo estratificado para el mismo tamaño muestral.
Sistemático	Simplicidad para extraer una muestra. Fácil de entender. Libre de error en la clasificación de sujetos. Facilidad para analizar e interpretar los resultados. Los sujetos no necesitan estar numerados.	Mayor error de muestreo que en el muestreo estratificado para el mismo tamaño muestral. Periodicidad en la lista de elementos de la población.
Proporcional estratificado	Permite comparaciones fáciles de los subgrupos. Normalmente más representativa que el sistemático o aleatorio simple. Se necesitan menos sujetos si los estratos están relacionados con la variable dependiente. Los resultados representan la población sin ponderar.	Requiere identificación de subgrupos de cada elemento de la población. Requiere conocimiento de la proporción de cada subgrupo en la población. Puede ser costoso y difícil preparar listas de los elementos de población en cada subgrupo.
No proporcional estratificado	Permite comparaciones fáciles de subgrupos. Normalmente más representativa que el sistemático o aleatorio simple. Se necesitan menos sujetos si los estratos están relacionados con la variable dependiente. Asegura el número adecuado de elementos en cada subgrupo.	Requiere identificación de subgrupos de cada elemento de la población. Puede ser costoso y difícil preparar listas de los elementos de población en cada subgrupo. Requiere la ponderación de subgrupos para representar la población.

(continúa)

(continuación)

Por grupos	Bajo coste. Eficaz con poblaciones numerosas. Permite el análisis de grupos individuales.	Menos exactitud que en el aleatorio, sistemático o estratificado. Puede ser difícil recoger datos de los elementos en un grupo. Requiere que cada elemento de la población se asigne sólo a un grupo.
No probabilística		
Por conveniencia	Menos lento y costoso. Facilidad de administración. Normalmente asegura un índice alto de participación. Se puede generalizar a sujetos similares.	Difícil de generalizar a otros sujetos. Menos representativa de una población identificada. Los resultados dependen de las características particulares de la muestra. Mayor probabilidad de error debido a los sesgos del experimentador o de los sujetos.
Intencionado	Menos lento y costoso. Facilidad de administración. Normalmente asegura un índice alto de participación. Se puede generalizar a sujetos similares. Asegura la recepción de la información necesaria.	Difícil de generalizar a otros sujetos. Menos representativa de una población identificada. Los resultados dependen de características particulares de la muestra. Mayor probabilidad de error debido a los sesgos del experimentador o de los sujetos.
Por cuotas	Menos lento y costoso. Facilidad de administración. Normalmente asegura un índice alto de participación. Se puede generalizar a sujetos similares. Tiende a proporcionar muestras más	Requiere información identificativa de cada sujeto. Difícil de generalizar a otros sujetos. Menos representativa de una población identificada. Los resultados dependen de las características

(continúa)

(continuación)

	representativas que el intencional o por conveniencia.	particulares de la muestra. Mayor probabilidad de error debido a los sesgos del experimentador o de los sujetos. Más lenta que el muestreo por conveniencia o el intencional.
--	--	---

2. **Hipótesis de investigación.** Si el investigador espera encontrar pequeñas diferencias o relaciones débiles, es deseable contar con una muestra tan numerosa como sea posible. Por ejemplo, el efecto de los cursos de formación en las puntuaciones de los tests estandarizados produce diferencias prácticas relativamente pequeñas, pero quizás, importantes. Este efecto, normalmente, no se detecta en estudios con un número poco numeroso de sujetos.
3. **Limitaciones financieras.** Obviamente el coste de realizar un estudio puede limitar el número de sujetos incluidos en la muestra. Es mejor estimar esos costes antes del comienzo del estudio.
4. **Importancia de los resultados.** En la investigación exploratoria es aceptable un tamaño muestral más pequeño porque el investigador está dispuesto a tolerar un mayor margen de error en los resultados. En investigación que obtenga como resultado el emplazamiento de niños en distintos programas, o cuando se gasta una gran cantidad de dinero, es imperativo para el investigador alcanzar una muestra lo suficientemente grande para minimizar el error.

TABLA 4.3: PORCENTAJE DE INGRESOS GASTADO EN ALQUILER

Clase de ingreso	Número de familias	Datos de población	Tamaño de la muestra			
			1 de 10	1 de 20	1 de 30	1 de 50
A	1748	26.5	26.6	25.9	28.3	27.1
B	2477	22.7	22.9	23.5	22.3	22.6
C	2514	19.8	18.1	17.2	17.2	18.0
D	1676	15.8	16.0	14.4	17.1	16.9
E	3740	11.3	11.0	10.1	11.2	11.5

Fuente: Rowntree, B. S. (1941), *Poverty and Progress: A Second Social Survey of York*. London: Longman, Green, reimpreso con permiso de Joseph Rowntree Charitable Trust.

5. **Número de variables estudiadas.** Se necesita una muestra mayor para los estudios que contengan muchas variables dependientes o independientes o para estudios en los que hay presentes muchas sin controlar.
6. **Métodos de recogida de datos.** Si los métodos de recogida de información no son altamente exactos o consistentes, se necesitará una muestra mayor para compensar el error intrínseco en la recogida de datos.
7. **Grado de exactitud necesario.** La exactitud de los resultados (el grado de confianza que puede darse a la afirmación de que los datos de la muestra son los mismos que los de la población) es mayor a medida que crece el tamaño de la muestra. Sin embargo, como demuestra el estudio de Rowntree, se alcanza un punto de rendimiento decreciente a medida que el tamaño muestral aumenta hasta un cierto porcentaje de la población.
8. **Tamaño de la población.** A medida que se incrementa el tamaño de la población, el investigador puede tomar progresivamente un porcentaje menor de sujetos de ésta.

En la investigación educativa una consideración importante a tener en cuenta con respecto al tamaño de la muestra es concluir que un estudio con una muestra relativamente pequeña que no encuentre diferencia o relación es verdadero. Por ejemplo, suponga que está estudiando la relación entre tipo de personalidad y capacidad de liderazgo y, con sólo veinticinco sujetos, encuentra que no existe relación. ¿Es razonable concluir que, en realidad, no existe tal relación? Probablemente no, ya que una razón importante para no encontrar relación es el reducido tamaño de la muestra. Es decir, que con un tamaño muestral mayor podría revelarse una relación. (Por otro lado, es posible informar de una relación «significativa» en un estudio que posea un número de sujetos muy numeroso, aunque la relación sea bastante reducida. Como veremos en los capítulos posteriores, esta relación «significativa», a menudo, se malinterpreta como una relación importante). Cuando los investigadores desean estar seguros de que el hecho de no hallar una relación o diferencia refleja exactamente la realidad, diseñan la investigación para asegurarse lo que se denomina «potencia» adecuada en el análisis. Una muestra de gran tamaño es el mejor método para incrementar la potencia de un estudio.

TÉCNICAS DE RECOGIDA DE DATOS

La investigación implica la recogida de información sobre las variables del estudio. El investigador elige a partir de una gran variedad de técnicas y de aproximaciones para la recogida de datos sobre los sujetos. Cada método tiene ventajas e inconvenientes y la aproximación específica adoptada debería

ser el mejor método para responder a la pregunta de investigación. A continuación presentamos algunos métodos habituales de recogida de información cuantitativa:

1. Cuestionarios
2. Entrevistas estructuradas
3. Tests
4. Observaciones estructuradas
5. Inventarios
6. Escalas de ordenación o clasificación
7. Puntuaciones adecuadas

Estos métodos serán expuestos con mayor detalle en los capítulos posteriores. Sin embargo, en este punto, es importante entender dos principios básicos de la medición, validez y fiabilidad, que son comunes a todos los métodos. El conocimiento de estos principios se emplea tanto para elegir instrumentos como para evaluar la adecuación de la recogida de datos presentada en los estudios de investigación.

Validez del test

La **validez del test** es el grado en que las conclusiones y empleos realizados, sobre la base de las puntuaciones obtenidas a partir de una prueba, son razonables y apropiados. La validez es un juicio de la idoneidad de una medida para las deducciones específicas, las decisiones, las consecuencias o los usos que se obtienen a partir de las puntuaciones que se generan. En otras palabras, validez es un concepto específico de la situación: la validez depende del propósito, de la población y de los factores circunstanciales en los que se realiza la medición. Los resultados de un test, de un cuestionario o de otro tipo de control pueden, por lo tanto, ser válidos en una situación y no válidos en otra.

Esta definición tiene implicaciones importantes para el diseño y para la evaluación de la investigación puesto que los resultados se relacionan directamente con la medida seleccionada. El investigador que está diseñando la investigación debe, primero definir claramente las consecuencias, los empleos o las decisiones que realizará a partir de los resultados. Debe, entonces, seleccionar un instrumento que proporcione una evidencia clara de que realizar tales deducciones o decisiones resultará provechoso.

La evidencia para la validez se describe en textos que resumen y critican las mediciones, como por ejemplo, los *Mental Measurement Yearbooks* (Anuarios de medición mental), en los manuales que describen información técnica sobre la medición y en los artículos de investigación en los que se maneja. Al evaluar

una investigación, el lector debería buscar los enunciados que se refieran a la validez y, luego, evaluar la evidencia proporcionada en relación con la conclusión o decisión tomada en el artículo. Por ejemplo, si se van a emplear datos normativos a partir de una prueba estandarizada, las normas deberían haber sido generadas a partir de una población que sea similar a los sujetos del estudio. En general, es importante tener en cuenta que las pruebas pueden ser válidas para algunos grupos y en algunas situaciones y no válidas para otros sujetos en otras situaciones.

Fiabilidad del test

La **fiabilidad de una prueba** hace referencia a la consistencia de su dimensión, el grado en el que las puntuaciones son similares en las diferentes formas de la misma prueba o en distintos momentos de la recogida de datos. A la hora de desarrollar puntuaciones fiables, la meta es minimizar la influencia de las puntuaciones en la situación concreta o de otras variables no relacionadas con el propósito de medida.

Los métodos específicos para la estimación y presentación de la fiabilidad son muy precisos y están explicados con más detalle en el capítulo 6. Tanto los que diseñan como los que leen una investigación deben interpretar la fiabilidad de la misma manera que la validez, y buscar pruebas de que se documenta la fiabilidad de las puntuaciones con garantía. Muchos estudios fallan al aceptar una hipótesis porque hay un error significativo en la medición de las variables (con puntuaciones más fiables, se podría aceptar la hipótesis).

Al leer el apartado de instrumentos/ pruebas de un estudio o diseñar la recogida de datos, se deben tener en cuenta algunos aspectos:

1. ¿Son las puntuaciones fiables para los sujetos de la investigación en particular?
2. ¿Son las características de los sujetos empleadas para establecer la validez y fiabilidad similares a las características de los sujetos en el estudio? Si no es así, ¿es razonable emplear estos instrumentos o pruebas?
3. ¿Se utilizan los mejores instrumentos? ¿Existen otros más fiables y que proporcionarían resultados más válidos?
4. ¿Por qué eligió esos instrumentos/ pruebas el investigador?
5. ¿Están descritas o citadas las pruebas suficientemente bien como para permitir que otro investigador repita la investigación?

El extracto 4.4 se refiere al apartado de instrumentos/ pruebas de un estudio. El nivel de descripción proporcionado es aproximadamente el esperado en la presentación de la mayoría de los tipos de investigación.

EXTRACTO 4.4**APARTADO DE INSTRUMENTOS****PREFERENCIA DE FUTURA DE CARRERA**

Las chicas marcaron la probabilidad de estudiar cada uno de los 19 empleos potenciales o categorías de carrera sobre una escala de 7 puntos que va desde «muy poco probable» hasta «muy probable». Cada categoría se presentaba con una etiqueta, seguida por una breve descripción (por ejemplo: servicio de comida, como camarero, camarera, cocinero, preparación de alimentos). Las categorías asociadas con la ciencia eran (a) paraprofesional de la salud (b) profesional de la salud con diplomatura (c) profesional científico o matemático con diplomatura (d) profesional de la salud con título superior y (e) profesional científico con título superior. Se proporcionaban ejemplos de cada categoría. Las puntuaciones para las cinco categorías se promediaron para crear una puntuación simple para una carrera de ciencia ($\alpha = .78$); las puntuaciones para las categorías (b) y (d) fueron promediadas para crear una puntuación de profesional de la salud ($\alpha = .81$); y se obtuvo el promedio de las categorías (c) y (e) para crear una puntuación de profesional científico físico ($\alpha = .77$).

Fuente: Jacobs, J. E., Finken, L. L., Griffin, N. L. y Wright, J. D. (1998), «The career plans of science-talented rural adolescent girls», *American Educational Research Journal*, 35 (4), 681-704.

Fuentes para localizar y evaluar las pruebas existentes

Una regla general de orientación en la realización de una investigación es elegir una prueba que tenga la fiabilidad y validez que el investigador necesita. Aunque la fiabilidad y la validez son las consideraciones más importantes en la selección de una prueba, existen otros, como los gastos de adquisición, disponibilidad, facilidad para su administración y corrección, limitaciones de derecho de autor, nivel de dificultad y adecuación normativa.

Aunque, a menudo, es difícil encontrar una prueba que cumpla todos los criterios que un investigador pueda necesitar, se han desarrollado miles de pruebas y es probable que alguna esté disponible para que pueda ser utilizada tal cual o modificada para cumplir un propósito concreto. La manera más fácil de localizar pruebas ya existentes es emplear fuentes que resuman la información sobre varias puntuaciones. Las fuentes de la siguiente lista son ampliamente utilizadas y son accesibles en la mayoría de las bibliotecas universitarias.

Test in Print, publicado periódicamente por el Instituto Buros (volumen 5, 1998, volumen 6, 2002), proporciona un resumen de tests revisados de todos los anuarios precedentes de medida. (Ediciones Earlier editado por Oscar Buros.)

- Handbook of Research Design and Social Measurement*, 5ª edición (1991), revisa y critica las puntuaciones más populares de ciencias sociales.
- Index to Tests Used In Educational Dissertations* (Fabiano, 1989) describe tests y poblaciones de tests usadas en tesis doctorales desde 1938 hasta 1980; ordenados por título y descriptores seleccionados.
- Directory of Unpublished Experimental Mental Measures*, volumen 6 (Goldman y Mitchell, 1995). Los referidos al periodo 1996-2000 están recogidos en el volumen 8, publicado en el año 2002, que describe más de 2.000 puntuaciones experimentales que no están disponibles comercialmente. Incluye referencias, fuente y objetivo en temas que van desde el ajuste educativo y la motivación hasta personalidad y percepción.
- The ETS Test Collection: El Educational Testing Service* (ETS) ha desarrollado varias fuentes que describen más de 10.000 tests y pruebas. El *Test Collection Bibliographies* recoge puntuaciones publicadas y no publicadas en distintas áreas, incluyendo rendimiento, actitudes e intereses, personalidad, poblaciones especiales y vocación/ocupación. Cada una de las 200 separatas describe pruebas y sus empleos apropiados y pueden solicitarse al ETS. Los *Tests in Microfiche* listan herramientas inéditas de investigación, también sobre una gran variedad de áreas. El *ETS Test Collection Catalog* incluye seis volúmenes: volumen 1: Test de rendimiento y dispositivos de medida (1993), volumen 2: Test vocacionales y dispositivos de medida (1995), volumen 3: Test para Poblaciones Especiales (1989), volumen 4: Test de Inteligencia y Aptitud Cognitiva (1990), volumen 5: Puntuaciones de actitud (1991) y volumen 6: Puntuaciones Afectivas y Test de Personalidad (1992). Cada volumen contiene descripciones completas de las pruebas, con índices por autor, título y sujeto. (<http://ets.org/testcoll>).
- Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business*, 4ª ed. (Maddox, 1996) (5ª edición, 2003), proporciona una descripción de más de 3.100 tests publicados, incluyendo propósito, coste, puntuación y editor.
- Test Critiques*, volúmenes 1-10 (Keyser y Sweetland, 1984-1994), ofrece revisiones minuciosas para pruebas de psicología, educación y empresa empleadas con frecuencia y recientemente publicadas y revisadas. Contiene información «orientada al usuario», incluyendo aplicaciones prácticas y usos, así como también aspectos técnicos y una crítica realizada por un especialista en psicometría. El complemento, *Test Critiques Compendium*, revisa los 60 tests principales de *Test Critiques* en un volumen.
- Mental Measurements Yearbooks* (MMY; Instituto Buros), proporciona revisiones de tests disponibles comercialmente en varias áreas, incluyendo carácter y personalidad, rendimiento, e inteligencia. Hace referencia a la mayoría de

los tests fomentando la investigación posterior. El MMY se ha publicado periódicamente durante seis años. El número 13 de MMY se publicó en 1998. El número 15 ha sido publicado en el año 2003. Pueden publicarse futuras ediciones de MMY cada dos años con suplementos editados en años alternativos. De este modo, si un test no se ha revisado, la información más reciente podría obtenerse consultando un volumen anterior. Conviene acceder a la web del Instituto Buros (<http://www.unl.edu/buros/>).

Test and Measurements in Child Development: Handbook I and II (Jonson, 1976), dos volúmenes describen cerca de 900 tests y pruebas inéditas para niños hasta los 18 años.

A Sourcebook for Mental Health Measures (Comrey, Backer y Glaser, 1973), esta referencia describe alrededor de 1.100 pruebas relacionadas con la salud mental, incluyendo delincuencia juvenil, personalidad y alcoholismo.

Measures for Psychological Assessment, una guía de 3.000 fuentes originales y sus aplicaciones (Chun, Cobb y French, 1974). Es una lista de tests usados en investigación y publicados en 26 publicaciones psicológicas diferentes.

Measures of Social Psychological Attitudes (Robinson y Shaver, 1973) describe cientos de pruebas de actitud empleadas en investigación psicológica y sociológica.

Socioemotional Measures for Pre-School and Kindergarten Children: A Handbook (Walker, 1973). Describe pruebas para medir: actitudes, personalidad, autoestima y niveles sociales de niños pequeños.

Handbook for Measurement and Evaluation in Early Childhood Education (Goddwin y Driscoll, 1980), un repaso general de puntuaciones afectivas, cognitivas y psicomotoras para niños pequeños.

Dictionary of Behavioral Assessment Techniques (Hersen y Bellack, 1988), proporciona descripciones de aproximadamente 300 pruebas que evalúan rasgos psicológicos y de comportamiento.

Desarrollo de pruebas

Aunque las pruebas disponibles son muchas, hay ocasiones en las que los investigadores tienen que desarrollar sus propias puntuaciones. La situación más común que requiere una medida desarrollada *ad hoc* es una investigación evaluativa para un ambiente específico. A no ser que la investigación tenga un importante impacto directo sobre los programas o individuos, no es normal que el investigador establezca sistemáticamente la fiabilidad y la validez (como se resume en el capítulo 6) antes de llevar a cabo el estudio. Una aproximación más habitual es desarrollar una prueba que parezca razonable y recopilar datos piloto sobre todo lo que se necesita revisar. Aunque, probablemente, no sea necesario

establecer sofisticadas estimaciones sobre fiabilidad y validez, es posible que una prueba posea una calidad tan baja que los resultados conseguidos no sean interpretables. Por lo tanto, si el investigador se enfrenta al desarrollo de una prueba, es importante que siga unos cuantos pasos básicos:

1. Familiarícese con las aproximaciones comunes para medir el rasgo o comportamiento de interés. Hay muchas fuentes existentes que resumen aproximaciones para la medida de variables como el rendimiento, las actitudes, los intereses, la personalidad ya autoestima.
2. Escriba los objetivos específicos para su prueba con un objetivo para cada rasgo o comportamiento de interés.
3. Después de leer sobre el área y cambiar impresiones con otros sobre qué aproximación sería la mejor para medir el rasgo, realice una tormenta de ideas con varios ítems para cada objetivo.
4. Pregunte a profesionales expertos en el área evaluada para revisar los ítems: ¿son claros?, ¿carentes de sesgos?, ¿concisos?, ¿es el significado el mismo para todos los lectores?
5. Localice una pequeña muestra de sujetos que sea similar a los que se emplearán en el estudio real y adminístreles la prueba. Esto podría denominarse un «test o ensayo piloto» de la prueba. Hay que comprobar la claridad, la ambigüedad en las frases, el tiempo para su cumplimentación, las instrucciones y cualquier problema que se haya experimentado.
6. Compruebe la distribución adecuada de las puntuaciones para cada apartado en la prueba. Si todas las respuestas a un ítem son las mismas, es difícil saber si la pregunta es inadecuada o si, de hecho, el rasgo carece de variabilidad. En tanto que las respuestas den como resultado una diversidad de puntuaciones, es probable que el ítem sea una medida adecuada del rasgo.
7. Revise, elimine y añada ítems donde considere necesario, dependiendo del *feedback* de la muestra del test piloto.

PROCEDIMIENTOS

En los estudios cuantitativos, el investigador planifica el procedimiento a utilizar para la recogida de datos y, en el caso de investigación experimental, la naturaleza y administración del tratamiento experimental. El investigador decide dónde se recogerán los datos (como por ejemplo, una escuela, una ciudad o en un entorno de laboratorio), cuándo se recogerán los datos (momento del día y época del año), cómo se recogerán los datos (por quién y de qué forma) y, si es necesario, detalles específicos del tratamiento experimental. Cualquiera de los procedimientos empleados para controlar el sesgo (como por ejemplo, compensar el

orden de las pruebas para controlar la fatiga o el aburrimiento del sujeto o asegurarse de que los observadores desconocen qué grupo recibe el tratamiento y cuál es el de control) se planifican e implementan como parte del procedimiento. En la presentación del estudio, el investigador debería exponer el procedimiento con el detalle suficiente como para permitir a otro investigador repetir el estudio.

VALIDEZ INTERNA DEL DISEÑO

La **validez interna**, que hemos definido como el grado de control sobre las variables externas, es más fuerte cuando el diseño del estudio (sujetos, pruebas y procedimiento) controla eficazmente las posibles fuentes de error para que éstas no se relacionen de un modo razonable con los resultados del estudio. Las fuentes de error para los estudios cuantitativos se interpretan como «riesgos», porque cada fuente puede invalidar los resultados del estudio.

Hay varias categorías o tipos de riesgos a la validez interna en la mayoría de los estudios cuantitativos. Cada una de esas amenazas, que es como se las denomina, se describe a continuación con ejemplos. Estas categorías han sido tomadas de Campbell y Stanley (1963), Cook y Campbell (1979) y McMillan (2000). Lo mejor es tener presente que los nombres de estas amenazas a la validez interna no deben interpretarse literalmente. A menudo, poseen un significado más amplio del que los nombres pueden sugerir en un principio. Aunque algunos de los nombres son específicos de este libro, la mayoría fueron establecidos hace muchos años por Campbell y Stanley (1963). Además, la mayoría de estas amenazas se concibieron originalmente para la investigación experimental. Aunque creamos que la mayoría de estas amenazas pueden relacionarse tanto con diseños experimentales como no experimentales, algunas sólo tienen sentido en el contexto de un experimento.

Historia

En el contexto de la validez interna, la **historia** hace referencia a los acontecimientos o incidentes extraños que suceden durante la investigación y afectan a los resultados. Esto es una amenaza para cualquier investigación que se realiza a lo largo de distintas etapas y puede llegar a convertirse en algo muy serio si se incrementa el tiempo entre medidas. Si durante el estudio ocurre algún hecho que se relacione de forma posible con la variable dependiente, resulta difícil saber si la variable independiente, el hecho en sí o una combinación de ambos produjeron el resultado. En este sentido, el hecho se «funde» con la variable independiente; ambos no pueden separarse.

La historia puede ocurrir «dentro» del estudio en tanto que los sujetos se ven afectados por algo que ocurre durante el tratamiento o «fuera» del entorno de la investigación. Por ejemplo, suponga que una clase estudia el Lejano Oriente y los investigadores están intentando determinar qué efectos tiene esta unidad didáctica sobre las actitudes multiculturales. Durante el estudio de la unidad tiene lugar una importante crisis en China. Si a los estudiantes les afecta la crisis, lo que a su vez influye en la forma en que responden al cuestionario sobre actitudes multiculturales, este hecho constituye una amenaza de historia a la validez interna del estudio. Las amenazas debidas a la historia también pueden intervenir dentro de un escenario de investigación. Por ejemplo, una serie de avisos inesperados, que distraen a una clase que recibe un método de instrucción determinado, afecta adversamente a la influencia de la enseñanza. Los alumnos de esta clase podrían puntuar más bajo que los de otras clases, pero el investigador no sabe si el resultado está causado por las distracciones o por el método de instrucción.

La historia, también, incluye variables de confusión que están asociadas con niveles de la variable independiente. Por ejemplo, imagine un experimento que compara dos métodos de enseñanza: un tratamiento se hace por la mañana y el otro por la tarde. La hora del día constituiría una amenaza a la historia asociada con los niveles de la variable independiente.

Selección

Hay que considerar dos tipos de amenazas de selección: aquéllas que tienen lugar en experimentos y en diseños *ex post facto* y las amenazas relacionadas con el muestreo. En los experimentos se forman grupos para estudiar una variable independiente de interés. Sin embargo, si hay una diferencia sistemática entre los grupos, es posible que los resultados puedan deberse a esas diferencias existentes. La amenaza de **selección** existe siempre que los grupos de sujetos no puedan asignarse de forma aleatoria y, aunque hay varias aproximaciones que ayudan a controlar este problema, el investigador siempre debería preocuparse de esta importante amenaza en los casos en los que la elección al azar no es deseable o resulta imposible (emparejamiento, comprobación de sujetos más de una vez [medidas repetidas], ajuste de las puntuaciones del postest en base a las características del grupo medidas inicialmente [análisis de covarianza] y administración a cada grupo de todos los tratamientos [contrarresta]).

Considere, por ejemplo, un profesor que quiere investigar si es mejor un enfoque de clase magistral inductiva para la enseñanza de adjetivos y verbos. El profesor consigue la colaboración de otra clase para realizar el estudio. Los dos profesores lanzan una moneda al aire para decidir cuál utilizará un

enfoque magistral y cuál el aprendizaje por descubrimiento. Los dos profesores evalúan el rendimiento administrando a cada grupo un pretest y un postest para determinar el aumento de conocimiento. Sin embargo, ocurre que la media de la puntuación CI del grupo de enseñanza magistral es 115 y la del grupo de aprendizaje por descubrimiento, 95. Aquí la selección es un problema fundamental, puesto que esperaríamos que el grupo con mayor CI rindiera más que el grupo con menor CI bajo casi cualquier condición. Tal amenaza a la validez interna podría inutilizar el estudio si no se controla y explica de algún modo. El profesor concluiría falsamente que el método de enseñanza magistral es más eficaz cuando en realidad su éxito aparente es debido a diferencias iniciales de capacidad.

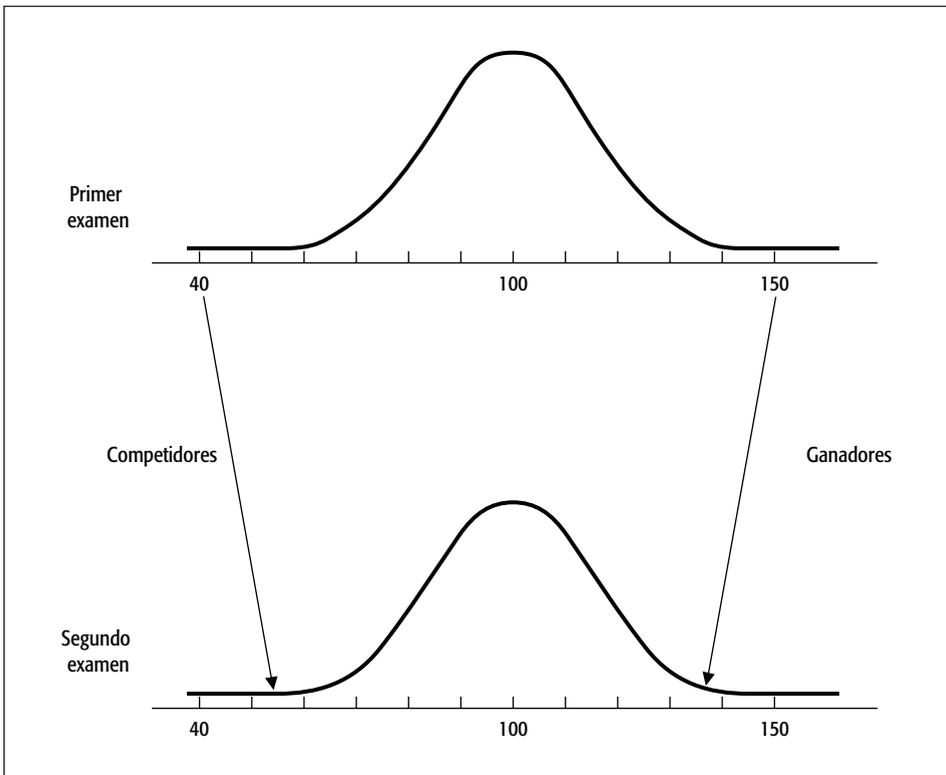
Como hemos mostrado anteriormente, la selección también está relacionada con la forma en que el investigador selecciona una muestra. Como apuntábamos, un problema común en investigación es el empleo de voluntarios para las muestras. El grupo de voluntarios puede estar más motivado o estarlo por razones especiales; de ahí que, podrían responder al tratamiento o a las preguntas de manera diferente a como lo haría un grupo de no voluntarios.

Regresión estadística

La **regresión estadística** hace referencia a la tendencia de los sujetos que puntúan muy alto o bajo en un pretest a puntuar cerca de la media («retroceder» hacia la media) en el postest, independientemente de los efectos del tratamiento. Todas nuestras puntuaciones tienen algún grado de error y la regresión estadística se produce debido a los cambios en el error del pretest al postest. Las puntuaciones en un postest serán diferentes de las del pretest para los estudiantes sobre la base de la probabilidad matemática en solitario a causa de este error. Para grupos de alumnos que puntúan tanto muy alto como muy bajo en un pretest, este error actúa para cambiar las puntuaciones en el postest de manera que están más cerca de la media del postest que la que estaban de la media del pretest.

Para ilustrar este concepto, piense en la figura 4.2 como una representación de la misma prueba realizada dos veces a dos grupos de alumnos en una clase en la que la puntuación media es 100. La puntuación media para los ganadores en la primera prueba fue 150 mientras que la puntuación para los competidores fue de 40. En la segunda prueba, esperaríamos que la puntuación media para los ganadores fuera más baja y la puntuación de los competidores mayor, incluso si su verdadera capacidad o su conocimiento permanecen igual.

FIGURA 4.2: ILUSTRACIÓN DE REGRESIÓN ESTADÍSTICA



La regresión es un problema cada vez que el investigador elige grupos intencionadamente por sus puntuaciones extremadamente altas o bajas. Por ejemplo, un distrito escolar quiere implantar un programa especial para mejorar la autoestima de los niños que puntúan bajo en un inventario de autoestima. Un análisis del impacto del programa podría examinar las puntuaciones después del programa (en un posttest), pero el investigador tendría que tener presente que incluso si no hay efecto del programa en absoluto, las puntuaciones bajas iniciales (en el pretest) mejorarían en algún grado debido a la regresión estadística. De forma similar, normalmente es difícil encontrar cambios positivos en programas para niños superdotados ya que, debido a la regresión, las puntuaciones del posttest tenderán a ser ligeramente más bajas por término medio.

Efectos del pretest

Siempre que la investigación utilice un pretest (alguna forma de medición que precede a un tratamiento o experiencia), es posible que éste ejerza por sí mismo

un impacto en los sujetos. El mero hecho de hacer un pretest puede proporcionar al sujeto práctica o motivación sobre el tipo de cuestiones preguntadas o familiarizar al sujeto con el material. Este tipo de **efecto del pretest** se encuentra en experimentos que miden el rendimiento en un corto espacio de tiempo y en investigación sobre actitudes o valores cuando se administra un pretest y un posttest a un único grupo. Si se utiliza como pretest un cuestionario de actitud, la simple lectura de las preguntas podría inducir al sujeto a pensar sobre el tema e, incluso, a cambiar sus actitudes. Un investigador podría, por ejemplo, estar interesado en estudiar el efecto de una serie de películas en las actitudes de los alumnos hacia los niños con discapacidad física. El investigador realiza un pretest, pone las películas y, luego, realiza un posttest para averiguar si ha habido cambios. Sin embargo, cualquiera de los cambios observados, puede haber estado causado por el pretest. Los ítems del cuestionario pueden haber sido suficientes para modificar las actitudes. El efecto del pretest no es una amenaza en diseños no experimentales.

Instrumentación

Existe una amenaza a la validez interna relacionada con la medición que se denomina instrumentación. La **instrumentación** hace referencia a que la forma en que los cambios en las pruebas o en las personas que se utilizan para recoger los datos pueden afectar a los resultados. Esta amenaza es particularmente seria en investigación observacional, en la que los observadores pueden llegar a estar cansados, aburridos o cambiar de alguna manera como para afectar al registro de los datos. Un buen ejemplo de cómo la instrumentación puede afectar a los resultados se observa cuando las puntuaciones del mismo test estandarizado se emplean para seguir la pista del rendimiento a lo largo de varios años. Si ha habido un cambio en las normas de la prueba y un nuevo formato, resulta problemático para comparar resultados entre el formato antiguo y el nuevo. El cambio de normas de 1995 y otros cambios en el SAT son un buen ejemplo de cómo esto puede llevar a errores de interpretación, puesto que el significado de la misma puntuación, por ejemplo, 500, es diferente en el nuevo formato. La medición es un cambio en el sujeto que resulta de realizar la prueba, mientras que la instrumentación es un cambio registrado en los resultados a partir de insuficiencias de la medición.

Mortalidad experimental

La **mortalidad experimental** (también llamada *mortalidad de sujetos*) ocurre en un estudio cuando los sujetos abandonan sistemáticamente o se pierden durante la investigación. Ésta es una amenaza para muchos estudios longitudinales

que duran más de varias semanas o meses. Por ejemplo, un estudio del efecto de un programa para ayudar a los alumnos de noveno curso con bajo rendimiento, realizado entre los cursos noveno y decimosegundo, tendría éxito, aparentemente, si los estudiantes con rendimiento más bajo abandonaran los estudios antes del 12º curso y no fueran incluidos en los análisis del postest. Para la mayoría de la investigación no experimental y de corta duración, la mortalidad no es una amenaza a no ser que el tratamiento sea específicamente exigente y provoque sistemáticamente que los sujetos de bajo rendimiento abandonen. En los estudios que sufren una pérdida diferencial de sujetos de diversos grupos debido a sesgos de selección o a la naturaleza de los tratamientos, la mortalidad es una seria amenaza a la validez interna. La mortalidad es básicamente el mismo problema que la selección, pero sucede después de que el estudio ya esté establecido y en marcha.

Maduración

La **maduración** hace referencia a los cambios en los sujetos de un estudio a través del tiempo que afectan a la variable dependiente. Los sujetos evolucionan y cambian como parte del crecimiento, y en la interpretación de la investigación que sucede durante un largo periodo de tiempo tales cambios deben ser considerados. Algunos cambios como, por ejemplo, tener hambre, cansancio, aburrimiento o enfado pueden ocurrir en un periodo de tiempo relativamente corto y se consideran también amenazas relativas a la maduración a la validez interna. Suponga que un investigador está estudiando las actitudes de los alumnos de cuarto curso hacia la lectura, las matemáticas y las ciencias. El investigador ha desarrollado una prueba y se la ha administrado a todos los sujetos en el mismo orden. Su realización es de media hora para la prueba de lectura y otra media hora para completar los ítems de matemáticas. ¿Cómo responderán a la prueba de ciencias? Probablemente estarán cansados, aburridos y faltos de atención; por esto, la maduración sería un problema importante si se emplean sus respuestas a los ítems de ciencias como indicativos de su actitud hacia las ciencias. Otros ejemplos se referirían a estudiantes que llegan a ser expertos debido a la experiencia, o alumnos de primer curso que aprenden a regatear un balón de baloncesto, no tanto por tener un profesor eficaz, como por madurar físicamente.

Difusión del tratamiento

En los diseños experimentales se administra un tratamiento a un grupo y el grupo alternativo o de control nunca entra en contacto con el del tratamiento experimental. Por ejemplo, si un psicólogo está estudiando la conducta de decir

mentiras y manipula los incentivos para mentir como variable independiente, un grupo podría recibir altos incentivos o recompensas por engañar y el otro, bajos incentivos, pero ningún grupo conocería el tratamiento que recibió el otro grupo. Sin embargo, si un profesor ha decidido comprobar este concepto entre los estudiantes de una clase y, a la mitad les dice que recibirán una gran recompensa y a la otra mitad, al mismo tiempo, que recibirán una pequeña recompensa, entonces cada grupo conoce las condiciones del otro. En tales circunstancias, los tratos se difunden entre todos los sujetos. Es posible que ambos tratamientos puedan afectar al otro grupo, como resultado de la **difusión del tratamiento**. La difusión también sucede si las consecuencias del tratamiento se extienden a los sujetos del grupo de control o comparación.

Efectos del experimentador

Los **efectos del experimentador** hacen referencia tanto a influencias deliberadas como a las no intencionadas que el experimentador posee acerca de los sujetos. Esto puede reflejarse en el hecho de tratar de forma distinta a los sujetos, por ejemplo, al emplear un tono de voz diferente, ser más apaciguador con un grupo que con otro, reforzar diferentes comportamientos, mostrar diferentes actitudes, observar de forma selectiva respuestas de diferentes sujetos y cualquier otro comportamiento que influya tanto en la conducta de los sujetos como en su evaluación por parte del investigador. Los efectos del experimentador también inciden si las características del investigador o de la persona que recoge los datos afecta a las respuestas de los sujetos, como por ejemplo, la vestimenta, la edad, el sexo, el nivel educativo y la raza. Por ejemplo, si un experimentador está llevando a cabo un estudio sobre la diferencia de conducta de alumnos de lo que llamamos profesores «expertos» comparada con la conducta de alumnos de profesores «novatos» y si los observadores registran las conductas de los alumnos de ambos profesores, es importante, que los observadores no sepan qué profesores son los expertos o los novatos. Si los observadores son conscientes del grupo en que están los alumnos (profesor experto o novato), este conocimiento puede influir en lo que los observadores descubren en las aulas. En la mayoría de las indagaciones que implica la aplicación de investigadores como una parte del estudio, lo mejor es no mencionarles los detalles específicos de la búsqueda. Necesitan tener sólo la información suficiente para poner en marcha la investigación con objetividad y recoger la información.

Repeticiones del tratamiento

En un experimento, se supone que el tratamiento debe repetirse hasta que cada uno de los miembros de un grupo lo recibe de la misma manera y de forma

separada e independiente de los otros miembros del grupo. De esta manera, si el investigador está poniendo a prueba un método nuevo de enseñanza en una clase, realmente sólo hay una respuesta al tratamiento; esto es, el tratamiento se realiza una vez. Cada clase es como un sujeto, de ahí que sean necesarias varias clases para realizar adecuadamente la investigación. Las **repeticiones del tratamiento** son una amenaza a la validez interna en el grado en que el número de sujetos en el estudio no sea el mismo que el número de repeticiones independientes del tratamiento.

Esta amenaza es una limitación especialmente problemática para la investigación educativa porque es muy difícil disponer de muchos grupos de estudiantes. A menudo, un estudio comparará dos tratamientos, uno administrado a una clase y el segundo a otra clase distinta. Aunque este tipo de diseño resulta normalmente de un tamaño muestral de dos, antes que del número de alumnos en las aulas, la amenaza de las repeticiones del tratamiento no significa que los resultados no sean interpretables. Más bien, lo que se debe reconocer es que los resultados, las interpretaciones y las conclusiones deberían hacerse con gran cuidado. Algunos tratamientos experimentales se administraron a los estudiantes como un grupo de asignación, pero la condición real del tratamiento es individualizada, como por ejemplo, un tipo determinado de deberes para casa. En estas circunstancias, las repeticiones del tratamiento no son una amenaza para la validez interna.

Efectos del sujeto

En una investigación ideal, los sujetos se comportan y responden de forma natural y honesta. Sin embargo, cuando las personas se implican en un estudio, a menudo, modifican su comportamiento simplemente porque entienden que son «sujetos» y, algunas veces, estos cambios afectan a los resultados. Los **efectos del sujeto** hacen referencia a las modificaciones de conducta del sujeto, admitidas por él mismo, como respuesta a la situación de la investigación. Si los sujetos tienen alguna idea del propósito del estudio o la motivación para hacerlo «bien», pueden alterar su comportamiento para responder más favorablemente. Los sujetos recogerán pistas del entorno y de las instrucciones, que les motivarán de maneras específicas. Estas pistas se llaman **características demandadas**. En la mayoría de los estudios, los sujetos quieren también presentarse a sí mismos de la manera más positiva posible. De este modo, pueden darse autopresentaciones positivas, conveniencias sociales o creencias de que se esperan determinadas respuestas, lo que podría afectar a los resultados. Por ejemplo, la mayoría de la gente quiere parecer inteligente, competente y emocionalmente estable, y pueden negarse a tratamientos que perciben como una manipulación negativa de sí mismos o pueden falsificar las respuestas para

parecer más positivos. Algunos sujetos pueden incrementar su comportamiento positivo o conveniente simplemente porque saben que están recibiendo un tratamiento especial (esto se denomina efecto *Hawthorne*, considerado por algunos investigadores como una amenaza a la validez externa; ver la sección sobre validez externa ecológica). Los sujetos del grupo de control pueden actuar con más empeño porque se ven en competición con un grupo de tratamiento o pueden estar motivados porque no recibieron el tratamiento (esto se denomina efecto *John Henry* o *rivalidad compensatoria*). Otros sujetos, cuando se dan cuenta de que no fueron seleccionados para el tratamiento que consideran es su preferido, pueden llegar a desmotivarse (*desmoralización por resentimiento*). Finalmente, muchos individuos reaccionarán positivamente, con una motivación o participación cada vez mayor, debido a que están haciendo algo nuevo y diferente (esto se denomina *efecto de la novedad*). Como puede ver, existen muchos efectos posibles del sujeto.

Conclusión estadística

En la mayoría de las investigaciones educativas, se emplea la estadística como una base para hacer conclusiones sobre presuntos efectos y relaciones. Hay varios principios estadísticos que si se vulneran pueden afectar a las deducciones realizadas a partir de los resultados, así como a las conclusiones subsiguientes de la investigación. Cook y Campbell (1979) enumeran varias amenazas en esta categoría que están más allá del alcance de este libro. Sin embargo, los investigadores necesitan entender que el empleo de estadísticos que suenan y parecen impresionantes no garantiza resultados válidos. Lo mejor es consultar con expertos en estadística familiarizados con estas amenazas.

VALIDEZ EXTERNA DEL DISEÑO

La **validez externa** hace referencia a la generalizabilidad de los resultados. Hay dos categorías generales de validez externa para diseños cuantitativos que necesitan ser consideradas al diseñar estudios o al evaluar los resultados de la investigación: validez externa poblacional y validez externa ecológica.

Validez externa poblacional

Los sujetos que se utilizan en una investigación poseen ciertas características y pueden ser descritos con respecto a variables como edad, raza, sexo y capacidad. Los resultados de un estudio, estrictamente hablando, sólo pueden generalizarse a otras personas que tengan las mismas, o por lo menos similares,

características que aquellos a los que se aplicó el experimento. El grado en el que los resultados pueden generalizarse a otras personas se denomina **validez externa poblacional**. Considere la situación predominante en muchas investigaciones psicológicas. Debido a limitaciones de tiempo, dinero y otros aspectos, los psicólogos recurren con frecuencia a estudiantes universitarios como sujetos de investigación. Estrictamente hablando, los resultados de tal investigación se limitan en generalizabilidad a otros estudiantes universitarios similares. En otras palabras, lo que podría ser cierto para ciertos estudiantes universitarios puede no serlo para estudiantes de sexto curso. De forma similar, la investigación realizada con alumnos de primaria no debería generalizarse a alumnos de secundaria, ni la realizada con hombres generalizarse a mujeres, ni la realizada con hispanos generalizarse a negros, y así sucesivamente. Un tratamiento podría ser efectivo con un tipo de alumno e ineficaz con otro. Si los sujetos para la investigación son voluntarios, los hallazgos pueden estar limitados a las características de los voluntarios.

Validez externa ecológica

La **validez externa ecológica** hace referencia a las condiciones de la investigación y al grado en el que la generalización de resultados se limita a condiciones similares. Las condiciones de la investigación incluyen factores como la naturaleza de las variables dependientes e independientes, entorno físico, hora del día o época del año, sensibilización al pretest y postest o los efectos causados por la presencia de un experimentador o un tratamiento. Entre estos factores se incluye el conocido **efecto Hawthorne**: la tendencia de las personas a actuar de forma diferente, simplemente, porque se dan cuenta de que son sujetos de una investigación. (Se llama efecto *Hawthorne* porque el estudio original fue realizado en la planta de Hawthorne de la *Western Electric* en Chicago. Aunque la validez del estudio original ha sido cuestionada, la etiqueta *efecto Hawthorne* perdura). Al igual que con los efectos del sujeto, las personas se ponen nerviosas, falsean respuestas para quedar bien o reaccionan de muchas otras formas debido a su conocimiento de aspectos de la investigación.

Debería apreciarse que muchos de los factores que limitan la generalizabilidad pueden actuar, también, como variables externas (amenazas a la validez interna) y afectan a la variable dependiente. A menudo se dice que tales factores se *confunden* con el tratamiento o estudio. Por ejemplo, si un estudio muestra un cambio en una clase matinal que ha recibido un método de aprendizaje y no se observa cambio en una clase vespertina que ha recibido un método diferente, la hora del día es una variable externa que podría afectar a la variable dependiente y está confundida con la eficacia del tratamiento. Puede ser que se obtengan resultados similares (esto es, generalizados) en otras clases matutinas. En este

caso, la generalizabilidad está limitada por el factor ecológico: hora del día; pero, la hora del día también es una amenaza a la validez interna. De forma similar, si un profesor está aplicando un currículum X y un profesor diferente un currículum Y, la variable profesor se confunde con la variable currículum y sería considerada una variable externa porque no es posible separar el efecto del profesor de los del currículum. La generalizabilidad de los resultados debe incluir los profesores así como también los currícula. Como hemos apuntado anteriormente, el efecto *Hawthorne* es un ejemplo de un factor que puede afectar tanto a la validez interna como a la externa.

Considere una vez más el psicólogo en un ambiente universitario. Los experimentos psicológicos son, a menudo, realizados en una habitación pequeña en la que el sujeto se encuentra solo. Este procedimiento es apropiado para controlar las amenazas a la validez interna, pero tiende a restringir el grado de una posible generalización, puesto que la forma en que responden los sujetos en un ambiente artificial puede no ser la misma que en condiciones naturales.

Un tipo especial de limitación a la generalizabilidad es el grado en que un tratamiento es una muestra representativa del constructo subyacente supuesto. Este tipo de validez externa fue denominada por Cook y Campbell (1979) **validez de constructo**. Se refiere a la adecuación con la que entendemos y comunicamos la complejidad del llamado paquete de tratamiento, particularmente, sus componentes que permitirán la respuesta del efecto observado como otra prueba del constructo subyacente. Por ejemplo, un programa específico de modificación de conducta puede emplearse como un método de aplicación de un principio del conductismo (modelado, refuerzo intermitente, saciedad, etc.); pero, para objetivos de generalización, el consumidor debe analizar la adecuación con la que el constructo está representado por el programa, y si se dan peculiaridades en el tratamiento que pueden dificultar la respuesta. Este problema se observa, en gran medida, en la investigación que estudia el impacto de un nuevo currículum o conjunto de materiales. Los resultados pueden ser impresionantes; pero, a menos que haya evidencia de que se ha usado un constructo generalizable y de que el tratamiento se entiende clara y fácilmente, la validez externa estará limitada.

Otra variación de la validez externa es tener cuidado en no concluir que lo que puede ser verdad para un grupo completo de sujetos también lo es para subgrupos. Por ejemplo, si la investigación en un instituto muestra una relación positiva entre la cantidad de tiempo empleada en los deberes y el rendimiento de todos los estudiantes, esto necesariamente no significa que esta misma relación se mantenga con alumnos de bajas o altas capacidades o que sea cierto tanto para los alumnos de segundo curso como para los del último año. Esto se denomina generalización *a lo largo de* una población y puede conducir a interpretaciones erróneas.

EXTRACTO 4.5 VALIDEZ EXTERNA

EXPOSICIÓN

Ya que sólo participaron un maestro y los alumnos de un aula, y ya que el estudio se centraba sólo en un currículum de arte, los resultados de este estudio deben ser aceptados experimentalmente hasta que sean respondidos. La coherencia de los resultados con la teoría existente y la investigación previa proporcionan alguna confirmación de su validez...

Aunque los resultados de este estudio y la investigación previa indican una superioridad pragmática de las estructuras de enseñanza cooperativa frente a la competitiva e individualizada, no se deduce que se deba abandonar toda individualización o competición. Se necesita una serie de estudios que demuestren las condiciones específicas bajo las que cada tipo de estructura de aprendizaje es útil y eficaz para lograr los resultados educativos deseados.

Fuente: Johnson, D. T., Jonson R. T., Jonson, J., y Anderson, D. (1976), «Effects of cooperative versus individualized instruction on student prosocial behavior, attitudes toward learning, and achievement», *Journal of Educational Psychology*, 68, 446-452.

Se puede ser tan estricto con respecto a la validez externa que, prácticamente, toda la investigación resulte útil sólo en casos concretos. Aunque es necesario tener en cuenta la validez externa de los estudios, es preciso ser razonables y no estrictos al interpretar los resultados. Por ejemplo, es habitual que los investigadores hagan alusión, en las secciones de discusión o conclusión del artículo, a las limitaciones de generalización de sus resultados. El extracto 4.5 es un buen ejemplo de investigadores que afrontan el problema de la validez externa a través de la limitación de la generalización que puede realizarse.

CONSIDERACIONES ÉTICAS Y LEGALES⁴

Dado que la mayor parte de la investigación educativa trata sobre seres humanos, es necesario entender las responsabilidades éticas y legales de llevar a cabo

⁴ Un análisis más detallado de consideraciones éticas y legales se resume en P. Reynolds (1982), *Ethics and Social Research*, Englewood Cliffs, NJ: Prentice-Hall; Homan, R. (1991), *The Ethics and Social Research*, Londres: Longman; American Psychological Association (2000) *The Ethics and Social Research with Human Participants*, Washington, DC: American Psychological Association; Howe, K. R. y Moses, M. S. (1999), *Ethics in Educational Research. In Review of Research in Education*, A. Iran-Nejad y P. D. Pearson (ed). Washington, DC: American Educational Research Association y directrices publicadas por American Educational Research Association.

una investigación. A menudo, los investigadores se enfrentan a situaciones en las que los costes potenciales del uso de métodos discutibles deben equilibrarse con los beneficios de la realización del estudio. Los métodos discutibles se producen debido a la naturaleza de las preguntas de investigación y a la metodología diseñada para proporcionar resultados válidos. Los costes incluyen daños o problemas psicológicos, como por ejemplo, ansiedad, vergüenza, pérdida de autoestima y ataques a la dignidad humana, o pueden incluir infracciones legales de los derechos humanos. Tales costes, como resultado potencial de la investigación, deben ser ponderados con los beneficios para el participante en la investigación como una mayor comprensión de sí mismo, satisfacción de ayudar y conocimiento de los métodos de investigación, así como otros beneficios más obvios para la teoría y el conocimiento del comportamiento humano.

Es responsabilidad última de cada investigador poner en la balanza estas consideraciones y realizar el mejor juicio profesional posible. Para hacerlo, es necesario que el investigador sea totalmente consciente de los principios éticos y legales que deben ser afrontados. Presentamos estos principios describiendo sus implicaciones.

Ética de la investigación

Generalmente se considera que la ética trata de las opiniones sobre lo que es correcto o erróneo, apropiado o inapropiado, bueno o malo. Naturalmente, existe cierto grado de desacuerdo sobre cómo definir lo que es éticamente correcto en investigación. Pero es una cuestión muy importante, de creciente interés para los ciudadanos, los investigadores y los legisladores. Muchos grupos de profesionales y de gobernantes han estudiado temas éticos con detenimiento y han publicado directrices para planificar y realizar investigación de forma que proteja los derechos y el bienestar de los individuos. Los principios publicados por la *American Educational Research Association* son los más relevantes para la investigación educativa. Otra fuente útil es la *American Psychological Association*⁵. Los principios de mayor interés para los educadores se presentan a continuación.

1. El investigador principal de un estudio es el responsable de los patrones éticos a los que se adhiere el estudio.
2. El investigador debe informar a los sujetos de todos los aspectos de la investigación que podrían afectar sobre el deseo de participar y contestar a

⁵ American Educational Research Association (1992), «Ethical standards of the American Educational Research Association», *Educational Researcher*, 21, 23-26; American Psychological Association (1992), «Ethical Principles of Psychologists and code of conduct», *American Psychologist*, 47, inserción especial, *Ethical principles in the conduct of research with human participants* (1982) Washington DC: American Psychological Association.

todas las preguntas de los sujetos acerca de las características que pueden tener efectos o consecuencias adversos.

3. El investigador debe ser tan sincero y honesto con los sujetos como sea posible. Normalmente, esto implica una divulgación completa del propósito de la investigación, pero existen circunstancias en las que ocultar información sobre la investigación o engañar a los sujetos puede estar justificado. Ocultar información significa que se revela a los participantes sólo una parte del propósito de la investigación. Esto puede hacerse en estudios en los que una divulgación completa afectaría seriamente a la validez de los resultados. Por ejemplo, en una investigación sobre las actitudes raciales de los estudiantes, puede ser suficiente informar a los alumnos de que el estudio indaga sobre las actitudes hacia los demás.

Un asunto más comprometido supone la investigación en la que, por decirlo claramente, el investigador miente deliberadamente a los sujetos. Un buen ejemplo es el clásico estudio sobre las expectativas del profesor de Rosenthal y Jacobson (1968). Los investigadores informaron a los profesores que ciertos alumnos habían sido identificados como «brillantes» en un test diseñado para predecir el progreso intelectual. De hecho, el «test» era una valoración de la inteligencia y los alumnos fueron identificados de forma aleatoria. En este diseño fue necesario decir a los profesores una mentira. ¿Está justificado tal engaño? Después de todo, en este caso los estudiantes sólo se beneficiarían de la falta de información y los resultados tuvieron implicaciones muy importantes.

Por un lado, se puede justificar el engaño sobre la base de la contribución a los resultados. Por otro lado, es una afrenta a la dignidad humana y al respeto por uno mismo y puede fomentar la desconfianza y el cinismo hacia los investigadores. Nos parece que sólo debería recurrirse al engaño en casos en los que (1) la importancia de los resultados potenciales sea mayor que los efectos perjudiciales de mentir; (2) la única manera válida de llevar a cabo el estudio sea mediante el engaño; y (3) se proporcione una información apropiada, en la que el investigador manifieste a los participantes la naturaleza y la razón del engaño tras la finalización del estudio. Engañar no significa que los sujetos no hayan tenido la oportunidad de elegir si participar o no en el estudio.

4. Los sujetos deben ser protegidos de malestar físico y mental, daño y peligro. Si cualquiera de estos riesgos fuera posible, el investigador debe informar a los sujetos.
5. La mayoría de los estudios requieren que el investigador obtenga el informe de consentimiento de los sujetos antes de que participen en la investigación. El **informe de consentimiento** se consigue proporcionando a los sujetos una explicación de la investigación, la oportunidad para

finalizar su participación en cualquier momento sin penalización y la divulgación completa de cualquier riesgo asociado al estudio. El consentimiento, normalmente, se consigue solicitando a los sujetos (o a los padres de los menores) que firmen un formulario que indica que han comprendido la investigación y su consentimiento para participar. Casi todas las recogidas de datos en colegios públicos que requieren la participación de los alumnos más allá de los exámenes normales necesitan los permisos de los padres, del distrito escolar y del director.

El consentimiento informado implica que los sujetos pueden decidir si participan. Sin embargo, existen circunstancias en las que parece aceptable que los sujetos nunca sepan que han participado. Algunas veces no resulta práctico o es imposible localizar a los sujetos; otras veces, el conocer su participación puede invalidar los resultados. Parte de la investigación educativa es poco indiscreta y no entraña riesgos para los sujetos (como por ejemplo, el empleo de datos de tests de estudiantes de los últimos diez años para apuntar tendencias de rendimiento). No obstante, el investigador infringe lo que muchos consideran que es el derecho ético de los participantes a tomar sus propias decisiones sobre la participación. En general, cuantos más inconvenientes o daños potenciales para los sujetos origine una investigación, tanto más grave es la cuestión ética de utilizarles como sujetos sin su consentimiento.

Es evidente que nunca se debería coaccionar a nadie a participar. La coacción se realiza en diferentes grados. En un extremo, los profesores pueden insistir a sus alumnos para que participen o los empresarios pueden «sugerir encarecidamente» a sus empleados que cooperen como sujetos. Se ejerce una persuasión sutil menos discutible convenciendo a los sujetos que están «beneficiando» a la ciencia, a un programa, o a una institución. El investigador puede indicar la libertad de elección para participar o no participar, pero el mensaje implícito «si no participas, nos decepcionas» también puede estar patente, por lo que sería una coacción parcial. En otros casos, simplemente se soborna a los sujetos. ¿Dónde termina la libertad de elección y empieza la coacción? A menudo es difícil saberlo, pero es responsabilidad del investigador ser consciente del poder de la coacción sutil y defender claramente la libertad del potencial participante para decidir si desea o no convertirse en un sujeto de la investigación. Siempre que sea posible, la participación debe ser voluntaria y debe minimizarse la invasión en la privacidad.

6. La información obtenida acerca de los sujetos debe ser confidencial, a menos de que, previamente, se acuerde lo contrario a través del consentimiento informado (ver sección de limitaciones legales). Esto significa que nadie tiene acceso a los datos individuales o a los nombres de los par-

ticipantes, excepto los investigadores, y que los sujetos sabían antes de participar quién vería los datos. La confidencialidad está salvaguardada si se asegura que los datos no pueden ser asociados a los sujetos individuales por el nombre. Esto puede lograrse de varias maneras (1) recogiendo los datos de forma anónima; (2) empleando un sistema para asociar los nombres a los datos que pueda destruirse; (3) recurriendo a una tercera persona para asociar los nombres a los datos y, luego, entregar los resultados al investigador sin los nombres; (4) solicitando a los sujetos que usen seudónimos o números; e (5) informando sólo de resultados grupales, no individuales. Boruch y Cecil (1979) proporcionan detalles de muchos procedimientos diferentes para asegurar la confidencialidad.

7. Para la investigación realizada a través de una institución, como por ejemplo, una universidad o un sistema educativo, es necesario obtener la aprobación para llevar a cabo la investigación antes de la recogida de los datos.
8. El investigador tiene la responsabilidad de considerar las posibles malas interpretaciones y usos inadecuados de la investigación y debería hacer todos los esfuerzos para comunicar los resultados de forma que se minimicen las malas interpretaciones.
9. El investigador tiene la responsabilidad de reconocer cuándo se ha privado de beneficios potenciales al grupo de control. En tales situaciones, la importancia de los resultados potenciales debería ser mayor que el daño potencial a algunos sujetos. Por ejemplo, un programa nuevo que pretenda aumentar el rendimiento en los niños con dificultades de aprendizaje puede ser ocultado a algunos niños con DA por opinar que se necesita un experimento para documentar la eficacia del programa. Durante el proceso, se ha denegado la participación en el programa a los controles que podían haberse beneficiado.
10. El investigador debe proporcionar a los sujetos la oportunidad de conocer los resultados del estudio en que están participando.

Cierto tipo de investigación, como por ejemplo, los estudios sobre los efectos de las drogas, obviamente tienen un peligro potencial que el investigador debe tener en cuenta cuidadosamente. Mucha de la investigación educativa no parece implicar ningún problema ético, pero el punto de vista del investigador puede estar sesgado y lo mejor es ser conservador y buscar el asesoramiento y la aprobación de otros. La mayoría de las universidades y agencias de financiación incluyen comités para la revisión de investigación con el objetivo de ayudar a asegurar la protección ética. Una de las continuas dificultades a las que se enfrentan los investigadores educativos es asegurar la cooperación de los directores, profesos-

res y padres. Esta dificultad queda reducida en los investigadores que tienen conciencia de los principios éticos y se adhieren a ellos.

Además hay una interesante, si no frustrante, interacción entre ser ético, por un lado y, por otro, diseñar la investigación para conseguir los datos mejores y más objetivos. Es relativamente sencillo, por ejemplo, observar el comportamiento de forma no perturbadora y es posible que los sujetos nunca sepan que formaron parte de un experimento. Como comentamos anteriormente, el efecto *Hawthorne* puede reducir la validez del estudio. Por lo tanto, para maximizar tanto la validez interna como la externa, lo mejor para los sujetos es no tener conocimiento de que están siendo estudiados. Suponga, por ejemplo, que un investigador pone a un cómplice en una clase para que registre discretamente la atención de los alumnos. ¿Tiene la obligación el investigador de contar a los alumnos que sus comportamientos están siendo registrados? Si los estudiantes son conscientes de que están siendo observados, ¿cambiará esto su comportamiento e invalidará los resultados? Tales situaciones presentan dilemas éticos y el investigador debe considerar los criterios citados con anterioridad para determinar la mejor línea de acción.

Limitaciones legales

La mayoría de las limitaciones impuestas a los investigadores desde 1974 se centran en la protección de los derechos y el bienestar de los sujetos. Estos requisitos, normalmente, son acordes con los principios éticos resumidos en el apartado anterior y están en constante estado de reinterpretación y de cambio por parte de los tribunales de justicia.

El decreto sobre los derechos educativos y la privacidad de la familia de 1974, conocido como la Enmienda Buckley, permite a los individuos conseguir acceso a la información que les incumbe, como por ejemplo, la puntuación en los tests, los comentarios del profesor y las recomendaciones. El decreto también dictamina que se necesita legalmente un permiso escrito de consentimiento para los datos que identifican a los estudiantes por su nombre. El consentimiento debe indicar la información que será desvelada, el propósito de la revelación y a quién será revelada. Se conceden excepciones a estos requisitos para la investigación que emplea registros escolares en los que los resultados son de «legítimo interés educativo» y se informa sólo de datos grupales. Debería tenerse en cuenta que los datos recogidos en un estudio, normalmente, pueden ser utilizados por los tribunales, incluso, si el investigador ha prometido confidencialidad a los participantes.

El *National Research Act* de 1974 requiere una revisión de la investigación propuesta por un grupo adecuado en una institución (división escolar o universi-

dad) para proteger los derechos y el bienestar de los sujetos. Aunque la mayor parte de las investigaciones sobre personas debe ser realizada por un grupo de estas características, existen algunas excepciones, como por ejemplo, investigar usando datos de tests que proceden de los programas normales de examen o analizando los datos públicos existentes, registros o documentos sin identificaciones individuales (ver Registro Federal 26 de enero de 1981, publicado por el Departamento de Salud y Servicios Sociales, para atender una lista completa de excepciones). Es recomendable que un grupo autorizado, antes que una sola persona, revise la investigación antes de su realización.

RESUMEN

Este capítulo presenta las características fundamentales del diseño de investigación cuantitativa. Hemos puesto una atención particular en la selección de sujetos y pruebas y sobre las variables que deberían ser tenidas en cuenta al diseñar e interpretar la investigación.

1. El diseño de investigación hace referencia a la forma en que se planifica y realiza un estudio.
2. El propósito de un buen diseño de investigación es aumentar la credibilidad de los resultados teniendo en cuenta tres fuentes de variabilidad: sistemática, error y externa.
3. Una investigación con validez interna controla las fuentes potenciales de sesgo o error.
4. La validez externa se refiere a la generalizabilidad de los resultados.
5. La elección al azar, el mantenimiento de factores constantes y los ajustes estadísticos pueden ser utilizados en estudios cuantitativos para mitigar las hipótesis rivales posibles.
6. El muestreo probabilístico se utiliza para poder hacer generalizaciones a una población más amplia.
7. El muestreo probabilístico se realiza a través de muestreo aleatorio simple, muestreo sistemático, muestreo aleatorio estratificado y muestreo por grupos.
8. El muestreo no probabilístico incluye el intencional, por conveniencia y por cuotas. Aunque tales muestras tienen gran disponibilidad y se pueden obtener fácilmente, su uso limita la generalizabilidad.
9. El tamaño muestral debería ser lo más numeroso posible sin llegar al punto en el que nuevos sujetos contribuyan con escasa o ninguna información.
10. Con el fin de obtener una validez y una fiabilidad aceptable para los sujetos que se utilizan en el estudio, deben escogerse cuidadosamente los

instrumentos de medida. La validez es una estimación de la idoneidad del uso de las puntuaciones y la fiabilidad es un indicativo de la consistencia de la evaluación.

11. Los investigadores deben intentar localizar pruebas existentes antes que desarrollar las suyas propias.
12. Las amenazas a la validez interna de los estudios cuantitativos incluyen selección, historia, regresión estadística, efecto del pretest, instrumentación, mortalidad experimental, maduración, difusión del tratamiento, efectos del experimentador, repeticiones del tratamiento, efectos del sujeto y conclusión estadística.
13. Las amenazas a la validez externa de los estudios cuantitativos se clasifican en características poblacionales y condiciones ecológicas.
14. La sección de procedimiento de un estudio debe describir cómo se recoge la información con el detalle suficiente como para permitir a otros investigadores rebatir o ampliar el estudio.
15. Los investigadores deben ser conscientes de las responsabilidades éticas y de las limitaciones legales que acompañan la recogida de datos y comunicación de la información.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. El propósito del diseño de investigación es...
 - a. seleccionar las pruebas para un estudio.
 - b. proporcionar una respuesta válida y creíble a un problema.
 - c. determinar el mejor tipo de análisis estadístico.
 - d. contrarrestar la validez interna y externa.
2. El sesgo en investigación...
 - a. responde a las preguntas de investigación.
 - b. proporciona control de las variables extrañas.
 - c. es un error que afecta a los resultados.
 - d. siempre está controlado.
3. En investigación cuantitativa, el control de las fuentes de error externas viene proporcionado por...
 - a. la elección aleatoria de sujetos.
 - b. el mantenimiento de las condiciones constantes.
 - c. el uso de factores como variables independientes.
 - d. las tres son ciertas.
4. La validez interna en el diseño de investigación cuantitativa es una valoración de...

- a. el control de variables extrañas.
 - b. la generalizabilidad de los resultados.
 - c. la manera de medir de un test, lo que intenta medir.
 - d. la exactitud del test.
5. Si ocurre un hecho no planificado mientras se lleva a cabo un experimento y afecta a los resultados, _____ es una amenaza a la interpretación de los resultados.
- a. la maduración
 - b. la regresión
 - c. el pretest
 - d. la historia
6. Los investigadores asignan sujetos a grupos de forma aleatoria principalmente para controlar la...
- a. selección.
 - b. maduración.
 - c. regresión.
 - d. mortalidad experimental.
7. La difusión del tratamiento ocurre cuando...
- a. se administran los tratamientos a todos los sujetos.
 - b. un grupo conoce y está influido por los tratamientos que reciben otros grupos.
 - c. el investigador hace un pretest del tratamiento.
 - d. un grupo se enfada con el otro.
8. La validez externa para la investigación cuantitativa hace referencia a...
- a. la generalizabilidad de los resultados.
 - b. la calidad de realización de la investigación.
 - c. las características de los sujetos.
 - d. el empleo de los resultados de la investigación sólo en el ambiente en el que se realizó la investigación.
9. Una muestra en investigación cuantitativa es todo lo siguiente excepto...
- a. el grupo seleccionado por el investigador.
 - b. el grupo de sujetos estudiados.
 - c. el grupo al que el investigador intenta generalizar los resultados.
 - d. el grupo seleccionado a partir de una población.
10. Una muestra aleatoria...
- a. debe ser lo más numerosa posible.
 - b. es aquélla en la que cada sujeto tiene la misma probabilidad de ser elegido.
 - c. es lo mismo que aleatorización.
 - d. la b y la c son correctas.
11. El muestreo aleatorio estratificado...

- a. utiliza condiciones físicas existentes para formar grupos.
 - b. toma muestras aleatorias de cada grupo.
 - c. es menos eficaz que el muestreo aleatorio simple.
 - d. necesita más sujetos que el muestreo aleatorio simple.
12. El grado en el que las puntuaciones de un test son significativas y apropiadas se llama...
- a. validez.
 - b. fiabilidad.
 - c. exactitud.
 - d. estabilidad.
13. La fiabilidad de un test es una estimación de...
- a. consistencia.
 - b. generalizabilidad.
 - c. qué tipo de error existe en el test.
 - d. la a y la c son correctas.
14. En la sección de procedimiento de la mayoría de los estudios cuantitativos, hay una indicación de...
- a. dónde, cuándo y cómo se recogieron los datos.
 - b. quién participa en el estudio.
 - c. cómo se analizaron los datos.
 - d. la a y la b son correctas.
15. Las muestras probabilísticas son...
- a. normalmente proporcionales.
 - b. usadas principalmente en investigación experimental.
 - c. obtenidas de una población más numerosa.
 - d. aproximadamente como las muestras de voluntarios.
16. El informe de consentimiento...
- a. se requiere para cualquier investigación.
 - b. proporciona a los sujetos una elección para implicarse en la investigación.
 - c. es aproximadamente lo mismo que informar a los sujetos.
 - d. un procedimiento que permite a los sujetos tener acceso a los resultados del estudio.

Problemas aplicados

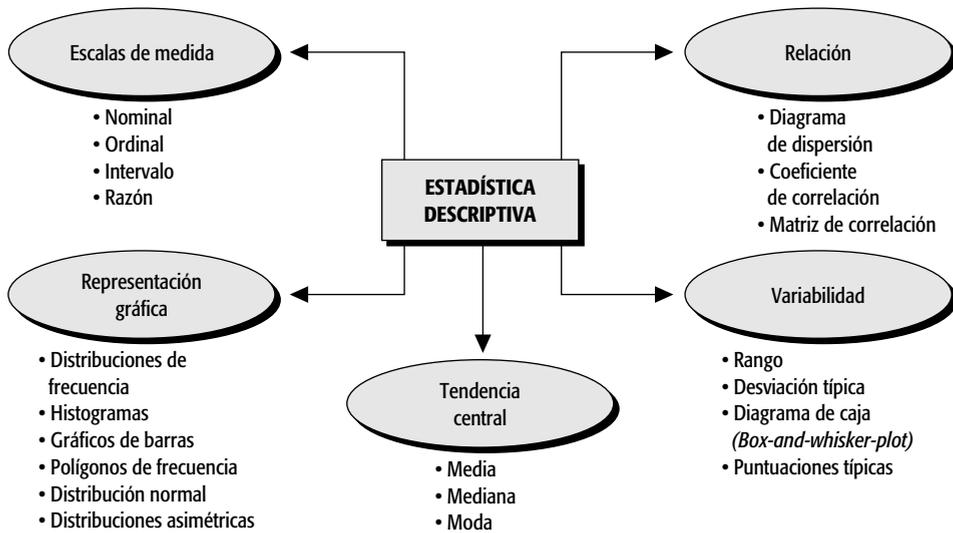
1. Para cada caso descrito aquí, enumere las amenazas potenciales a la validez interna y externa.
- a. Dos investigadores diseñaron un estudio para estudiar si el rendimiento en educación física está afectado por pertenecer a una clase en que todos los alumnos tienen el mismo sexo o a una clase con alumnos de ambos

sexos. Consiguieron la cooperación de un profesor. Se ofrecieron tres grupos de la misma clase de tenis, un grupo masculino, un femenino y uno mixto. Los investigadores controlaron la variable instructor utilizando la misma persona como instructor para cada grupo, informándole sobre el estudio y subrayándole la necesidad de mantener las mismas actividades educativas para cada grupo. Se planteó un grupo por la mañana, otro a mediodía y el tercero por la tarde. Los estudiantes se matriculaban en este curso siguiendo el mismo procedimiento que para el resto de los cursos, aunque había una nota a pie de página acerca de la composición de género en la planificación de los cursos. Se realizó un pretest para controlar las diferencias existentes entre los grupos.

- b. En este estudio se examina el efecto de la guardería en la conducta de sociabilidad de los niños. Un grupo de padres voluntarios accedió a participar en el estudio (los investigadores pagan parte de los gastos de la guardería). Los niños fueron asignados al azar a partir de un colectivo de voluntarios para asistir o no a una guardería de su elección. Los observadores midieron el nivel de sociabilidad antes y después de nueve meses de guardería, observando a los niños en el patio.
- c. Un inspector desea conocer si una emisión de bonos será aceptada en una próxima elección. Los listados de registros de contribuyentes de bienes raíces se obtienen de la oficina del condado. De esta lista se llama por teléfono, dos semanas antes del día de las elecciones, a una muestra al azar del 10% de 3.000 personas y se les pregunta si tienen intención de votar sí o no.
- d. La junta de dirección de la escuela de *Green County* decidió que quería una evaluación de las actitudes hacia las ciencias de los alumnos de noveno curso. Se diseñó un cuestionario y se distribuyó en enero a todos los profesores de ciencias de dicho curso. Se le dijo a cada profesor que administrara el cuestionario en seis semanas, calculara las puntuaciones medias para cada pregunta y devolviese los cuestionarios y los resultados a la oficina del distrito. Se les pidió a los profesores que emplearan el tiempo de una sola clase para minimizar la interferencia con el curso. Se devolvieron el sesenta por ciento de los cuestionarios.

CAPÍTULO 5

Estadística descriptiva



PALABRAS CLAVE

estadística	asimetría
estadística descriptiva	asimetría positiva
estadística inferencial	asimetría negativa
escalas de medida	medidas de variabilidad
nominal	rango
ordinal	desviación típica
intervalo	rango percentil
razón	varianza
distribución de frecuencias	diagrama de caja
histograma	<i>(Box-and-whisker-plot)</i>
gráfico de barras	puntuaciones típicas
polígono de frecuencia	puntuaciones z
medidas de tendencia central	diagrama de dispersión
media	relación positiva
mediana	relación negativa
moda	coeficiente de correlación
distribución normal	matriz de correlaciones

Este capítulo explica algunos procedimientos estadísticos comunes que se utilizan para describir la validez y fiabilidad de una prueba y para presentar los datos en un estudio cuantitativo. Los procedimientos representan maneras detalladas de resumir y de organizar los datos que se han recogido. Se presentan los procedimientos estadísticos descriptivos empleados más frecuentemente (aquéllos que describen fenómenos) incluyendo: distribuciones de frecuencia, medidas de tendencia central, medidas de variabilidad y medidas de relación. Se discuten las razones para el empleo de cada procedimiento con ejemplos de artículos publicados para ilustrar cómo se presentan los datos.

INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

La investigación cuantitativa se basa, en gran medida, en números para presentar los datos, realizar el muestreo y proporcionar estimaciones de validez y fiabilidad de la prueba. Los números normalmente están acompañados por palabras extrañas e irreconocibles y por símbolos, incluso más extraños, y se manipulan por medio de algo denominado *estadística*. Como si fuese magia, la estadística lleva a conclusiones. A menudo los lectores de investigación prefieren simplemente saltarse cualquier cosa relacionada con la estadística. En palabras de un especialista prominente en medición educativa: «para la mayoría de los educadores, la mera

contemplación del término «estadística» evoca imágenes similares a una plaga bubónica y la abolición del trabajo indefinido» (Popham, 1981, p. 66).

Incluso, cuando a algunos estadísticos puede gustarles la imagen que acabamos de describir, en realidad, los conceptos y principios fundamentales de la estadística son fácilmente comprensibles. Para comprender la estadística, no es necesario tener un cierto nivel en matemáticas y no hay que memorizar fórmulas complejas. De hecho, aprender estadística puede resultar divertido, especialmente considerando la gran cantidad de palabras nuevas que se aprenden que son perfectas para impresionar a los amigos y a la familia.

Bromas aparte, existen importantes razones para que todos los educadores consigan un dominio funcional de los principios estadísticos:

1. Entender y criticar artículos profesionales (por ejemplo, ¿eran correctas las herramientas estadísticas empleadas?).
2. Mejorar la evaluación del aprendizaje de los alumnos.
3. Dirigir, incluso de forma modesta e informal, estudios de investigación (por ejemplo, ¿cómo deberían analizarse los resultados?).
4. Entender evaluaciones de programas, de personal y de sistemas.
5. Ayudar a estar mejor preparado como ciudadano y consumidor, tomando decisiones basadas en datos o argumentos cuantitativos.
6. Mejorar la profesión educativa proporcionando habilidades uniformes para comunicar, debatir y discutir una investigación que tenga implicaciones para la práctica educativa.

Tipos de estadística

Las **estadísticas** son métodos de organizar y analizar datos cuantitativos. Estos métodos son herramientas diseñadas para ayudar al investigador a organizar e interpretar números derivados de la medición de un rasgo o variable. La mera presencia de procedimientos estadísticos no nos asegura una calidad alta en una investigación. Mientras la contribución de algunos resultados depende de la aplicación del procedimiento estadístico adecuado, la calidad de la investigación depende más de la correcta conceptualización, del diseño, de la selección de sujetos, pruebas y procedimientos. La estadística es un lenguaje internacional que sólo maneja números. Estadística y números no se interpretan a sí mismos y el significado de la estadística se deriva del diseño de investigación. Por supuesto, el empleo inadecuado de la estadística invalida la investigación, pero la interpretación de los resultados estadísticos depende de un diseño muy cuidadoso de realización del estudio, esto es, depende, en gran medida, de la producción de datos cuantitativos de alta calidad.

Hay dos amplias categorías de técnicas estadísticas: descriptiva e inferencial. La **estadística descriptiva** transforma un conjunto de números u observaciones en índices que describen o caracterizan los datos. La estadística descriptiva (a veces

llamada *estadística de resumen*) se usa para resumir, organizar y reducir grandes cantidades de observaciones. Frecuentemente, la reducción da como resultado un rango de números, que derivan de fórmulas matemáticas para representar todas las observaciones sobre cada grupo de interés. La estadística descriptiva representa y se enfoca en *qué es* con respecto a los datos de muestra, por ejemplo: «¿cuál es la nota media en lectura de los alumnos de quinto curso?, ¿cuántos profesores encuentran valioso su trabajo?, ¿qué porcentaje de estudiantes quiere ir a la universidad? y ¿cuál es la relación entre el nivel socioeconómico de los niños y la efectividad de los refuerzos simbólicos?». El uso de estadística descriptiva es la forma más conveniente de resumir los datos y resulta indispensable para interpretar los resultados de una investigación cuantitativa.

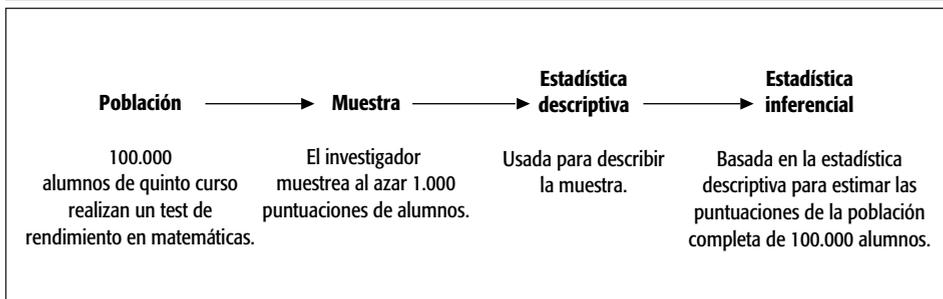
La **estadística inferencial**, por otro lado, se emplea para hacer deducciones o predicciones sobre la similitud de una muestra con la población de la que se ha extraído la muestra. La estadística inferencial se emplea con mucha frecuencia para presentar datos, puesto que muchas cuestiones de investigación requieren la estimación de las características de la población a partir de una muestra disponible de sujetos o comportamiento. El capítulo 9 trata con mucho detalle la función y tipos de estadística inferencial. La estadística inferencial depende de la estadística descriptiva. Por lo tanto, sin una comprensión completa de la estadística descriptiva, la estadística inferencial tiene muy poco sentido. La figura 5.1 ilustra la relación entre la estadística descriptiva y la inferencial. Muestra cómo podría un investigador primero tomar una muestra de una población, emplear estadística descriptiva para describir la muestra y, después, estadística inferencial para estimar el valor verdadero de la puntuación del test para la población.

Hay muchos tipos de estadística descriptiva que los investigadores pueden elegir para caracterizar un conjunto de datos. La opción depende, normalmente, de dos factores: el tipo de escala de medida empleado y el propósito de la investigación. Habitualmente se indica la escala de medida y, como se describe en la siguiente sección, existen técnicas estadísticas descriptivas que se corresponden con cada escala. El propósito de la investigación o el tema de investigación dependen, en realidad, de un conocimiento de las diferentes técnicas estadísticas ya que cada técnica ofrece información para dar respuesta a determinados tipos de preguntas. Por ello, aquí presentamos cada una de las técnicas descriptivas con ejemplos de los temas de investigación a los que van dirigidas.

Escalas de medida

Medir en educación implica, normalmente, la asignación de números a cosas para diferenciar una cosa de otra. Sin embargo, a diferencia de la medida de fenómenos físicos como el peso, la densidad o la longitud, los investigadores pueden usar números de maneras diferentes para la investigación de problemas.

FIGURA 5.1: RELACIÓN DE ESTADÍSTICA DESCRIPTIVA E INFERENCIAL



Estas formas diferentes están basadas en cuatro propiedades de los números. Las cuatro propiedades son: los números pueden ser distintos de otro (por ejemplo, 10 es diferente a 13; 0 es diferente a -5); los números están en relación con otros (por ejemplo, 13 es mayor que 10; -3 menor que 0); los números pueden relacionarse con cada uno de los otros en unidades identificadas (por ejemplo, 10 es cinco unidades de 2 y mayor que 5); y los números puede relacionarse proporcionalmente (por ejemplo, 10 es el doble que 5; 25 es a 5 como 30 es a 6). Estas propiedades, a su vez, determinan lo que los expertos en psicometría mencionan como **escalas de medida** o niveles de medida. Hay cuatro escalas de medida: nominal, ordinal, intervalo y razón. Estos términos, a menudo, se emplean para describir la naturaleza de la medida indicando, por ejemplo, que se usó una «medida nominal» o «medición nominal».

Nominal. El primer y más rudimentario nivel de medida se llama **nominal**, *categorico* o *clasificatorio*. La palabra *nominal* implica *nombre*, que describe lo que consigue esta escala, una denominación de categorías de gente mutuamente excluyentes, sucesos u otros fenómenos. Ejemplos comunes de niveles nominales incluyen la clasificación sobre la base de: color de ojos, género, afiliación política y tipo de grupo de lectura. Los grupos reciben un nombre para poder ser diferenciados, pero esto no supone ningún orden (un grupo no va antes o después que otro) y no hay indicación de la manera en que los grupos difieren unos de otros. Los investigadores a menudo asignan cifras a los distintos grupos (por ejemplo, sí = 1, no = 2, quizás = 3), pero esto es sólo para una codificación apropiada para el análisis de datos. Los datos nominales dan como resultado variables categóricas y los resultados se presentan como frecuencias en cada categoría.

Ordinal. El segundo tipo de escala de medida se denomina **ordinal** y, como su propio nombre indica, la medición de este tipo asume que las categorías de la variable teóricamente pueden ser ordenadas por rango desde la mayor hasta la menor. De esta forma, cada valor puede relacionarse con los demás como igual a,

mayor que, o menor que. En otras palabras, hay un orden inherente a las categorías. Ejemplos de medición ordinal son la clasificación de los miembros de una clase a través de la media de sus puntuaciones, la clasificación de ideas desde la más importante hasta la menos importante y el uso de rangos percentiles en los tests de rendimiento.

Intervalo. Las medidas de intervalo comparten las características de las escalas ordinales y, además, indican intervalos iguales entre cada categoría. Las escalas de intervalo dan significado a la diferencia entre números proporcionando una unidad constante de medición. La diferencia o intervalo entre 5 y 6, por ejemplo, es la misma que entre 18 y 19. Las puntuaciones percentiles asociadas a la curva normal, por ejemplo, no son un intervalo porque la distancia entre puntos percentiles varía dependiendo de los percentiles comparados. Hay una diferencia mayor entre los percentiles extremos (por ejemplo, 2 y 3 ó 95 y 96) que entre los percentiles cercanos a la mitad de la distribución. Ejemplos de escalas de intervalo incluyen: temperaturas Fahrenheit y Celsius y la mayoría de los tests estandarizados.

Razón. Las escalas de razón representan el tipo más refinado de medición. Las escalas de razón son ordinales y de intervalo y, además, los números pueden compararse mediante razones: esto es, un número puede compararse significativamente diciendo que es dos o tres veces otro número, o un medio o un cuarto de otro número. Son ejemplos de mediciones en escala de razón la distancia alcanzada, la fuerza expresada como peso levantado o las vueltas en una carrera. Sin embargo, la mayoría de mediciones en la educación no se expresan en forma de razón. Los educadores piensan en términos de menor que o mayor que, no en múltiplos (por ejemplo, un estudiante es más cooperativo o menos cooperativo, no el doble de cooperativo o la mitad de cooperativo).

Aunque no siempre es fácil identificar la escala de medida de algunas variables, es importante distinguir entre el nivel nominal y otros «más altos». El manejo de muchos de los procedimientos estadísticos más comunes, como por ejemplo, la media y la varianza, requiere una escala de medida de intervalo o de razón, aunque una escala ordinal a menudo es aceptable. La elección de otros procedimientos estadísticos más avanzados depende de si los datos son nominales o están en los niveles más altos. Si, por ejemplo, un investigador quiere comparar los estudiantes pertenecientes a una minoría y los que no, sobre la base de sus elecciones de carreras, los datos son nominales, y ciertos procedimientos estadísticos serían apropiados para analizar los datos. Si, por otro lado, estos mismos estudiantes fueran comparados por su rendimiento o por sus actitudes hacia la escuela, sería apropiado un conjunto diferente de procedimientos estadísticos porque la escala de los datos rendimiento y actitud es ordinal o de intervalo. Estas diferencias se seguirán tratando en el capítulo 9.

FIGURA 5.2: ESCALAS DE MEDIDA

Escala	Características de la escala	Ejemplos
RAZÓN	Los números representan cantidades iguales desde el cero absoluto. Las puntuaciones pueden compararse como razón o porcentaje.	Edad, dólares, tiempo, velocidad, tamaño de clase
INTERVALO	Las diferencias iguales entre números representan diferencias iguales en la variable o atributo que se mide.	Año (a-C.), ° F, ° C
ORDINAL	Los números representan orden de clasificación de la variable que se mide.	Cualquier variable clasificada, percentiles, clase social
NOMINAL	Los números distinguen entre las categorías. Los números no representan cantidad o grado. La asignación de números a los grupos es arbitraria.	Sexo, etnia, partido político, tipo de personalidad

Fuente: Glass, G. V., y Hopkins, K. D. (1996), *Statistical Methods in Education and Psychology* (3ª ed.) Needham Heights, MA: Allyn y Bacon.

REPRESENTACIÓN GRÁFICA DE DATOS

Cuando se recogen los datos, las observaciones deben organizarse para que el investigador pueda interpretar los resultados fácil y correctamente. Esta sección presenta tres métodos comunes para representar datos grupales.

Distribución de frecuencia: un dibujo de un grupo

En la mayoría de los estudios hay muchas puntuaciones diferentes y si estas puntuaciones se ordenan sin tener en cuenta sus valores, como en la tabla 5.1, es difícil darle sentido a los datos. La organización más simple de las puntuaciones sería listarlas de mayor a menor y crear lo que se llama una distribución por orden de importancia. La distribución por orden de importancia se transforma en una distribución de frecuencia por la indicación del número de veces que se alcanza cada puntuación, como se indica en la tabla 5.2.

También es frecuente combinar puntuaciones en intervalos de clase y contar el número de puntuaciones en cada intervalo, como se indica en la tabla 5.3. Los intervalos son especialmente útiles para los datos en los que algunos números son iguales (como por ejemplo, la clasificación de estados por su renta *per cápita*). Debido a que las distribuciones de frecuencia agrupadas sólo indican el número

total de puntuaciones dentro de cada intervalo, ocasionalmente se usa un gráfico de tallo y hojas para mostrar las puntuaciones reales en el intervalo. Los «tallos» son los intervalos de puntuación. Normalmente se listan verticalmente y para cada uno se incluye una fila que muestra las puntuaciones.

TABLA 5.1: PUNTUACIONES DE EXAMEN DESORGANIZADAS DE CINCUENTA ESTUDIANTES

47	37	41	50	45
39	49	44	43	40
42	43	42	46	40
44	45	47	45	45
36	45	46	48	44
42	48	40	43	37
46	45	45	44	42
43	43	42	43	41
44	45	42	44	36
44	38	44	46	42

TABLA 5.2: DISTRIBUCIÓN DE FRECUENCIAS DE LAS PUNTUACIONES EN TABLA 5.1

Puntuaciones en orden de importancia	Cuentas	Frecuencia (f)
50	1	1
49	1	1
48	11	2
47	11	2
46	1111	4
45	1111 111	8
44	1111 111	8
43	1111 1	6
42	1111 11	7
41	11	2
40	111	3
39	1	1
38	1	1
37	11	2
36	11	2
		n = 50

Las distribuciones de frecuencias son muy útiles para responder muchas preguntas importantes. Indican rápidamente las frecuencias mayor y menor de puntuaciones; la forma general de la distribución (por ejemplo, grupos de puntuaciones en ciertos lugares o puntuaciones dispersas de un modo uniforme) y si algunas puntuaciones están aisladas de las demás.

Histogramas y gráficos de barras

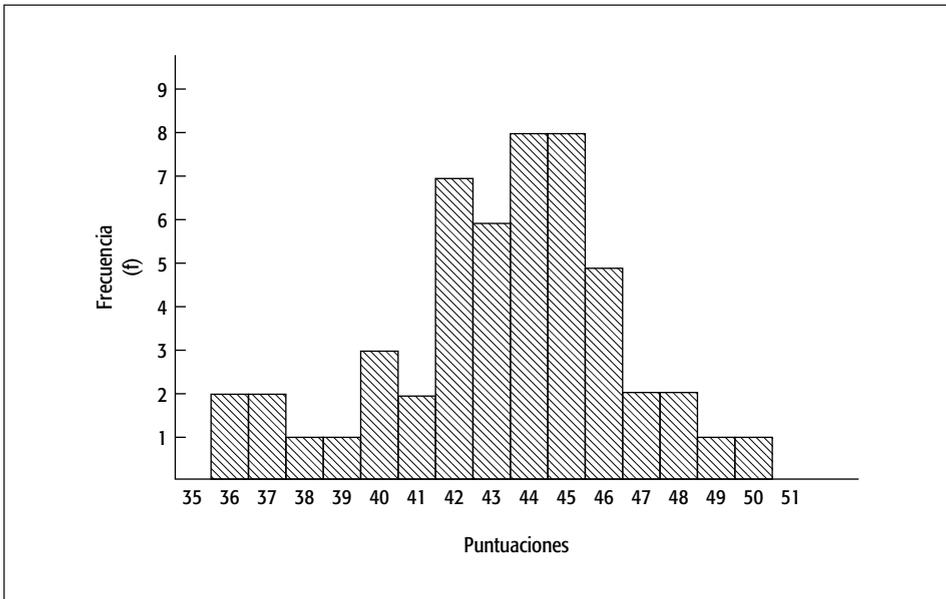
Los datos de frecuencia se muestran de forma más efectiva gráficamente. Un tipo de ilustración utiliza columnas en un gráfico de dos dimensiones para representar la frecuencia de cada puntuación o intervalo. Esta forma de representar una distribución de frecuencia se llama **histograma**. Los datos de las tablas 5.1 y 5.2 se representan como un histograma en la figura 5.3. En este ejemplo, la dimensión vertical del gráfico lista las frecuencias de las puntuaciones y la dimensión horizontal el orden de importancia de las puntuaciones de menor a mayor. Las columnas se dibujan en el gráfico de forma que se correspondan con los resultados. De la misma manera, el extracto 5.1 muestra cómo los histogramas pueden describir los resultados de un artículo.

TABLA 5.3: DISTRIBUCIÓN DEL INTERVALO DE FRECUENCIA DE LAS PUNTUACIONES DE LA TABLA 5.1

Intervalo	Frecuencia (f)
48-50	4
45-47	14
42-44	21
39-41	6
36-38	5
	n = 50

Un **gráfico de barras** se parece mucho a un histograma, con columnas que representan una imagen de los resultados. En un gráfico de barras, sin embargo, la ordenación de las columnas es arbitraria, mientras que en un histograma hay un orden de menor a mayor. Los gráficos de barras se usan, por tanto, con variables nominales como, por ejemplo, género, estado, afiliación política y variables categóricas similares que no tienen orden implícito. Se ilustra un gráfico de barras en el extracto 5.2.

FIGURA 5.3: HISTOGRAMA DE PUNTUACIÓN A PARTIR DE LA TABLA 5.1



Los histogramas son eficaces porque proporcionan una imagen fácilmente comprensible de los resultados. Sin embargo, la imagen puede distorsionarse manipulando el espacio de los números a lo largo del eje vertical del gráfico. Los intervalos entre las frecuencias de puntuación pueden variar y el tamaño de las unidades que se utilizan puede cambiarse para dar diferentes imágenes. Por ejemplo, un investigador astuto puede hacer que una pequeña diferencia parezca grande incrementando el espacio entre unidades de medida. Considere los dos gráficos de la figura 5.4. Cada gráfico tiene los mismos datos, pero los resultados visuales son diferentes.

Polígonos de frecuencia

Otra forma de ilustrar una distribución de frecuencia es usar un **polígono de frecuencia**. Un polígono de frecuencia es muy similar a un gráfico de barras excepto que los puntos en vez de barras están dibujados y conectados por una línea. La figura 5.5 muestra nuestro ejemplo de datos en un polígono de frecuencia. Observe que esta representación es muy similar a la figura 5.3.

Finalmente, también es útil representar gráficamente la distribución curvando las líneas rectas de un polígono de frecuencia. La famosa curva normal, discutida más adelante en este capítulo, es un ejemplo del uso de esta técnica.

FIGURA 5.4: GRÁFICOS DE PUNTUACIONES DE LECTURA DE ALUMNOS DE PRIMER, SEGUNDO Y TERCER CURSO

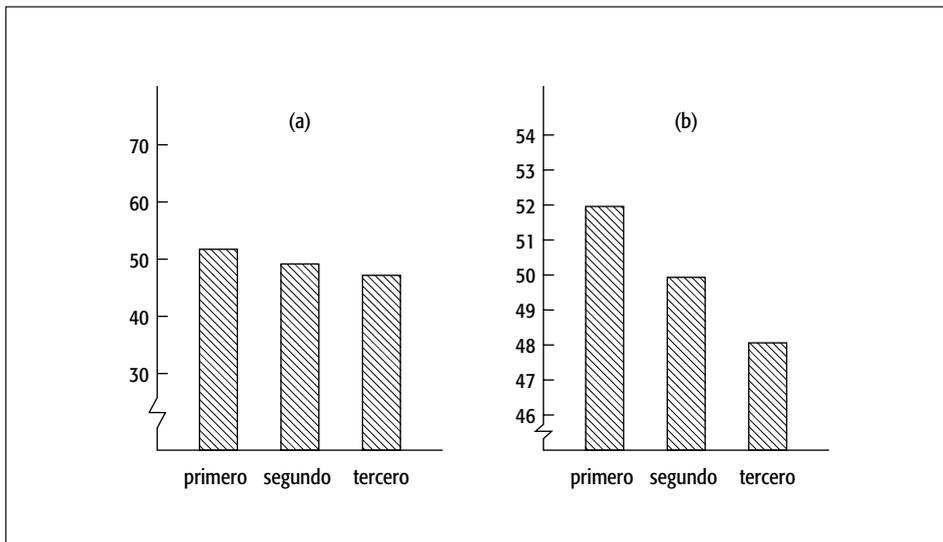
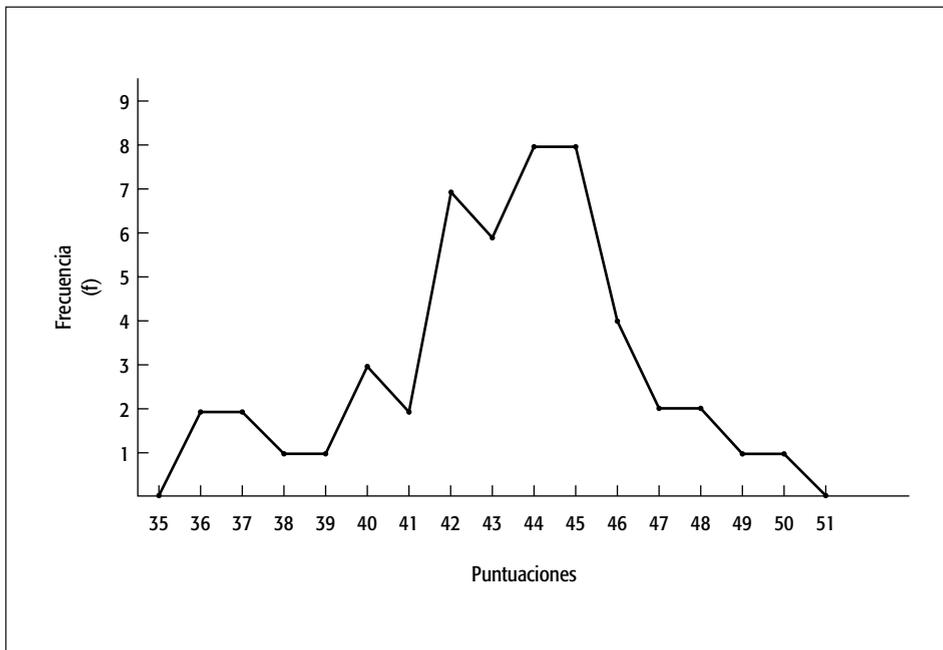


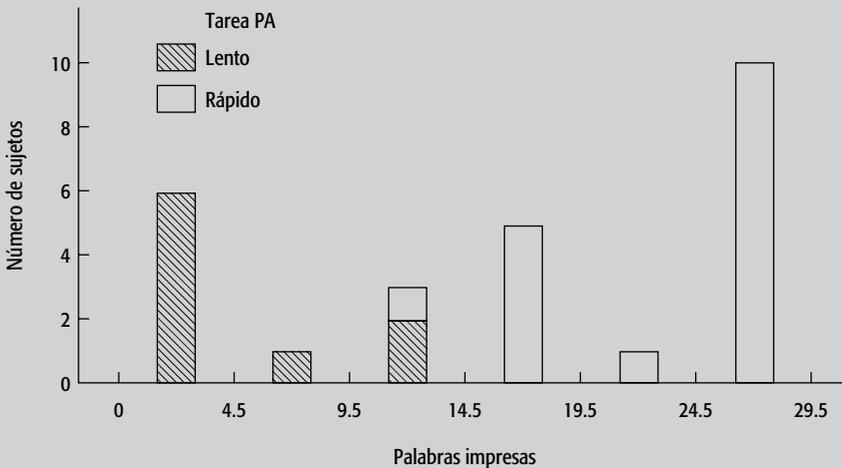
FIGURA 5.5: POLÍGONO DE FRECUENCIA DE PUNTUACIONES DE LA TABLA 5.1



EXTRACTO 5.1 HISTOGRAMA

En la figura 1 se presenta un histograma de la distribución de las puntuaciones de reconocimiento de palabras. Los sujetos se dividieron en dos grupos. Aquéllos que aprendían los sonidos en menos de 10 intentos se representan en blanco, aquéllos que necesitaban 10 o más intentos, en negro. A partir de esta figura, resulta evidente que hay poca coincidencia entre los dos grupos de sujetos. Aquéllos con extenso repertorio de palabras escritas aprendían los sonidos fácilmente. Los que tenían un repertorio pequeño, no.

FIGURA 1: Distribución de rápido y lento aprendices de sonidos en la tarea de la identificación de palabra impresa en el experimento 1. (PA 5 aprendizaje de pares asociados)



Fuente: Ehri, L. C. y Wilce, L. S., «The mnemonic value of orthography among beginning readers», *Journal of Educational Psychology*, 71, 26-40. Copyright «1979 de la American Psychological Association. Reimpreso con permiso de Línea Carlson Ehri.

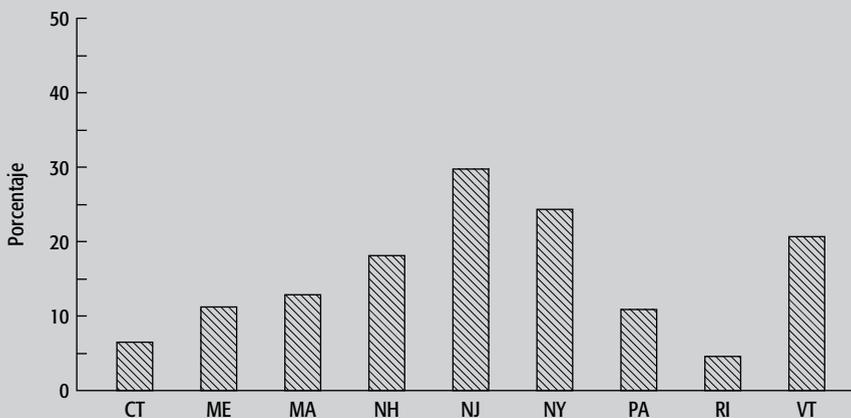
MEDIDAS DE TENDENCIA CENTRAL

Para la mayoría de conjuntos de datos, resulta útil tener alguna idea de la puntuación u observación media o típica además de conocer la distribución de frecuencia. Aunque la palabra *media* posee muchas connotaciones, en investigación sólo la media hace referencia a la puntuación promedio. Otros dos índices, la mediana y la moda, también proporcionan información sobre puntuaciones típicas de un grupo. Juntos, estos tres índices se denominan **medidas de tendencia**

central. Cada uno proporciona un índice numérico de la puntuación típica de una distribución. (Se pueden encontrar los cálculos para éstos y otros estadísticos descriptivos en el Apéndice B).

EXTRACTO 5.2 GRÁFICO DE BARRAS

Porcentaje de cambio en la matriculación en la escuela pública de 9 a 12 años, por estado: el Nordeste, 1996 a 2000



Fuente: *Projections of Education Statistics to 2002*, (1991), Washington, DC: National Center for Education Statistics.

La media

La **media** (simbolizada por \bar{X} o M) es simplemente la media aritmética de todas las puntuaciones. Se calcula sumando todas las puntuaciones y dividiendo la suma entre el número de puntuaciones. Si, por ejemplo, tenemos una distribución de 5, 8, 9 y 2, la media es 6 ($5 + 8 + 9 + 2 = 24$; $24/4 = 6$). La media es la medida de tendencia central más frecuentemente empleada porque se usan todas las puntuaciones para su cálculo. La desventaja de la media es que cuando una distribución contiene puntuaciones extremadamente altas o bajas, las más atípicas de la distribución, la media se desplaza hacia la puntuación extrema. Si, por ejemplo, una distribución contiene las puntuaciones de 4, 5, 7 y 40, la media sería 14. Puesto que en este caso la mayoría de las puntuaciones son considerablemente menores que 14, la media es un poco engañosa con respecto a la tendencia central.

EXTRACTO 5.3**LA MEDIA**

En la tabla 3 se presentan las medias para los tres periodos de recogida de datos. Muestran un descenso en el número y porcentaje de elecciones de amistad interracial en la clase de control entre el pretest y el posttest y un incremento durante este periodo en la clase experimental.

TABLA 3: Medidas de seguimiento sociométrico

Medida	Número medio de elecciones de amistad					
	Control ^a			Experimental ^b		
	Pretest	Posttest	Seguimiento	Pretest	Posttest	Seguimiento
Misma raza	5.95	6.37	7.35	4.00	3.77	4.00
Otra raza	2.65	1.58	.80	1.62	2.73	2.44
Porcentaje	30.8	19.9	9.8	28.8	37.2	37.9

^an = 20. ^bn = 16.

Fuente: Slavin, R. E. (1979), «Effects of biracial learnings teamson cross-racial friendships», *Journal of Educational Psychology*, 71. Reimpreso con permiso del autor.

La media se emplea habitualmente cuando se presentan datos de investigación cuantitativa y es esencial para la interpretación de resultados en los que se comparan unos grupos con otros. El extracto 5.3 ilustra el uso de la media en un artículo.

La mediana

La **mediana** es ese punto que divide una distribución ordenada en mitades que tienen un número igual de puntuaciones. Por esto, el cincuenta por ciento de las puntuaciones está por debajo de la mediana y el otro cincuenta por ciento se encuentra por encima de ésta. La mediana no está afectada por los valores reales de las puntuaciones. Por ejemplo, la mediana del conjunto de puntuaciones 10, 15, 16, 19 y 105 es 16, debido a que la mitad de las puntuaciones están por encima de 16 y la otra mitad por debajo. Dieciséis sería de esta manera un indicador de tendencia central mejor que la media, que es 33. Si una distribución contiene un número par de puntuaciones, la mediana es el punto medio entre las dos puntuaciones centrales (por ejemplo, para las puntuaciones 2, 2, 4, 7, 8 y 12, la mediana es 5,5).

La mediana se usa para describir datos que pueden tener puntuaciones extremas, como por ejemplo el nivel de ingresos en los Estados Unidos. Las medianas también se emplean, a veces, para dividir un grupo de entrevistados en dos grupos iguales. Un investigador puede, por ejemplo, obtener un indicador del grado de éxito percibido por cada entrevistado en una escala de 7 puntos (éxito extremo = 7, fracaso extremo = 1). Si el investigador quisiera dividir el grupo de sujetos entre los que tienen una alta y baja percepción del éxito, puede utilizar la mediana. Este procedimiento se denomina *división por la mediana*. El extracto 5.4 muestra cómo puede utilizarse la mediana en la presentación de una investigación.

EXTRACTO 5.4
LA MEDIANA

Una tendencia entre las familias que tiene claras implicaciones en las necesidades educativas dentro de las comunidades urbanas es el descenso de los recursos financieros que todas las familias están experimentando ahora, con un impacto incluso mayor entre las familias urbanas de minoría étnica. El último informe de la mediana de ingresos nacionales de familias blancas fue 18.370\$, para hispanos 12.570\$ y para negros 10.880\$ (ACYF, 1980). Mientras el 16% de todos los niños está por debajo del nivel de pobreza, un niño negro tiene 4 veces más posibilidades de estar por debajo de este nivel, el 11% de los blancos y el 42% de los negros viven en la pobreza (Edelman, 1980).

Las madres solteras tenían una mediana de ingresos que estaba muy por debajo del total y de la que ganan las familias con dos padres; las madres negras tenían una mediana de ingreso que era sólo el 40% del de las familias con dos padres, las hispanas tenían el 39% y las madres blancas tenían el 38%. La menor proporción de madres blancas se debe al hecho de que casi la mitad de las madres blancas no trabaja, a pesar del marcado incremento del empleo urbano para blancos (ver tabla 1).

TABLA 1: Ingreso medio y porcentaje de grupos étnicos desempleados y por debajo de la línea de la pobreza

Año		Negros	Hispanos	Blancos
1977	Familia de marido y mujer	13,832	13,432	18,756
	Madres solteras	5,598	5,247	8,799
1978		10,880	12,570	18,370

Fuente: Mc Adoo H. P., (1981), «Youth, school, and the family in transition», *Urban Education*, 16, 261-277. Copyright 1981 de Urban Education. Reimpreso con permiso de Sage Publications, Inc.

La moda

La **moda** es simplemente la puntuación que aparece más frecuentemente en una distribución. La moda es un índice bruto de tendencia central y raramente se utiliza en investigación educativa. Tiene utilidad sólo cuando interesa conocer la puntuación u observación más frecuente o cuando los datos están en formato nominal. La palabra *moda* se utiliza, quizá más frecuentemente, para describir una distribución indicando que la distribución es bimodal (dos modas) o trimodal (tres modas). Estos términos se usan aun cuando sólo exista una moda pero existen, al menos, dos puntuaciones que tienen frecuencias definitivamente más altas que el resto.

Relaciones entre las medidas de tendencia central

Mientras que una distribución de puntuaciones sea relativamente simétrica, la media, la mediana y la moda serán aproximadamente iguales. En lo que se refiere a una **distribución normal**, estos índices son exactamente los mismos. La distribución normal (ver figura 5.11 en la página 202) origina una curva simétrica en forma de campana. La curva normal es la distribución teórica que se usa para transformar datos y calcular muchas estadísticas. Aunque muchas variables educativas (por ejemplo, numerosas cifras de puntuaciones de rendimiento) se distribuyen siguiendo una curva normal, los datos de un estudio específico pueden disponerse desigualmente; esto es, las distribuciones son asimétricas y las puntuaciones tienden a agruparse en un extremo de la distribución o en el otro. Tales distribuciones se llaman **asimétricas** y con las distribuciones asimétricas la elección de medidas de tendencia central se convierte en algo muy importante. Las distribuciones tienen una **asimetría positiva** si la mayoría de las puntuaciones se sitúa en el límite inferior de la distribución con pocas puntuaciones altas y son **negativamente asimétricas** si la mayoría de las puntuaciones está localizada en el límite superior. Para recordar la diferencia entre asimetría positiva y negativa, piense en la forma curvada de la distribución formando una flecha o un puntero. Si forma una flecha que apunta en dirección positiva o hacia arriba, la distribución es positivamente asimétrica y si es en dirección negativa o hacia abajo, la distribución es negativamente asimétrica. Esto es, usando la mediana o la moda como punto de referencia, la media es mayor en una distribución simétrica positiva y menor en una distribución simétrica negativa. En realidad, se puede considerar que la media es positiva o negativamente asimétrica en relación con la mediana. En la figura 5.6 se ilustran distribuciones positiva y negativamente asimétricas con sus correspondientes medias, medianas y modas. Observe que la media en cada distribución está más alejada de la cola de la distribución que la mediana o la moda y que ésta última es la más lejana.

Para ilustrar con más detalle esta relación, considere el siguiente ejemplo. Suponga que un profesor quiere presentar una puntuación media de lectura para su clase. Tiene una puntuación de lectura de cada uno de los 20 alumnos, con un rango entre 5 y 80. La distribución de puntuaciones se representa en la tabla 5.4.

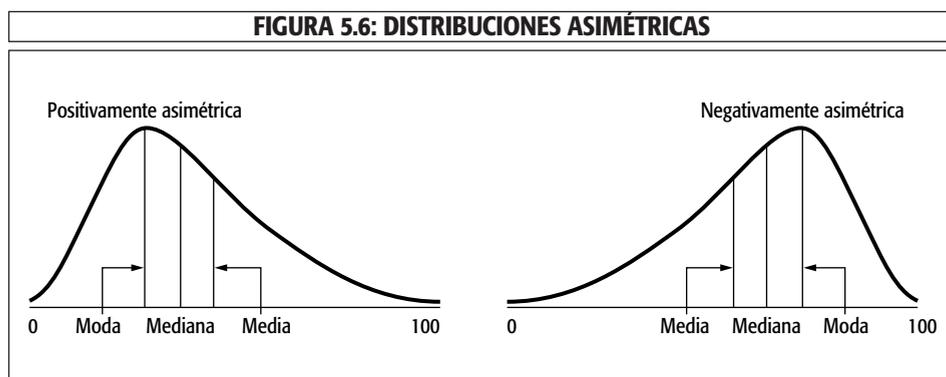


TABLA 5.4: DISTRIBUCIÓN DE FRECUENCIAS DE PUNTUACIONES DE LECTURA

Puntuaciones	(f)
5	8
10	4
12	2
15	2
80	4
	n = 20

Si el profesor presenta el promedio como la media, sería 22,7. La mediana es 10 y la moda 5. ¿Cuál es correcta? Debido a unos pocos alumnos que puntuaron muy bien (80), la distribución es positivamente asimétrica y, en consecuencia, la mediana es probablemente el indicador simple más exacto. Sin embargo, en tales casos, es probablemente mejor presentar la media para los alumnos que puntuaron entre 5 y 15 (8,4) y presentar las cuatro puntuaciones más altas por separado, o bien, presentar tanto la media como la mediana. Dado que en educación muchas distribuciones son por lo menos un poco asimétricas, a menudo, lo mejor es presentar tanto la media como la mediana.

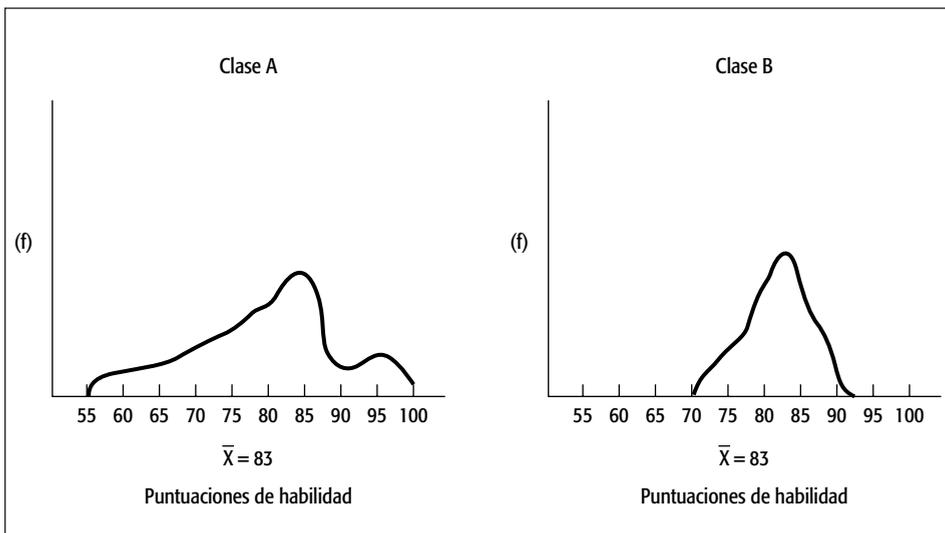
MEDIDAS DE VARIABILIDAD

La tendencia central sólo es un índice que se utiliza para representar un grupo de puntuaciones. Para proporcionar una descripción completa, se necesita una segunda medida estadística. Esta estadística hace referencia a una medida de variabilidad. Las **medidas de variabilidad** muestran cómo está de separada la distribución de puntuaciones de la media de la distribución o cuánto, sobre el promedio, las puntuaciones difieren de la media. Las medidas de variabilidad también se denominan en términos generales medidas de dispersión o distribución.

La necesidad de las medidas de dispersión se ilustra en la figura 5.7. Esta figura muestra cómo dos aulas con la misma puntuación media pueden ser muy diferentes. En la clase B los alumnos son más bien homogéneos, similares unos a otros, con pocos alumnos con alto o bajo rendimiento. Sin embargo, en la clase A, el profesor tiene un rango amplio de rendimiento, un grupo heterogéneo de alumnos cuyas puntuaciones se distribuyen entre 55 y 100.

O suponga que una persona va a apostar un sábado a un partido de baloncesto entre los Bombers y los Dunkers. La sección de deportes del periódico no tiene la estadística de los jugadores individuales, pero el periodista deportivo dice que los jugadores de ambos equipos tienen aproximadamente la misma altura: el promedio de altura esté en 2,02 y 2,05 metros, respectivamente, para los Bombers y los Dunkers. Con sólo la media como ayuda para tomar una decisión, el apostante realiza una apuesta por los Dunkers. Cuando ve el programa con las alturas de los jugadores, descubre una limitación de la media.

FIGURA 5.7: DISPERSIÓN DE PUNTUACIÓN



Bombers	Dunkers
Leary, defensa: 6'0"	Regen, defensa: 6'5"
Burns, defensa: 6'3"	Lambiotte, defensa: 6'6"
Parker, delantero: 6'5"	Hambrick, delantero: 6'8"
Gallagher, delantero: 6'7"	Lang, delantero: 6'9"
Robinson, central: 7'3"	Wergin, central: 6'10"
$\bar{X} = 6'6\frac{1}{2}"$	$\bar{X} = 6'7\frac{1}{2}"$

Como el juego de los Bombers consiste en sacar ventaja de la altura de Robinson sobre la de Wergin para marcar, el apostante se da cuenta de que la media falla al informar sobre las características de la distribución. Los Dunkers tienen una variabilidad pequeña, mientras los Bombers tienen una variabilidad alta, por lo que ¡el apostante pierde la apuesta!

La variabilidad, por lo tanto, nos habla de la diferencia de la distribución entre las puntuaciones. Aun cuando podamos usar palabras como *alto*, *bajo*, *grande*, *pequeño* y *mucho* para describir el grado de variabilidad, es necesario poseer índices más precisos. Las dos medidas más comunes de variabilidad son el rango y la desviación típica.

El rango

El **rango** es la medida de dispersión más obvia. Simplemente es la diferencia entre las puntuaciones mayor y menor de la distribución. Si, por ejemplo, la menor de las 30 puntuaciones en un test es 65 y la mayor 90, el rango sería 25 ($90 - 65 = 25$). Debido a que sólo hay dos puntuaciones involucradas en el cálculo del rango, es muy fácil obtenerlo. Sin embargo, también es una medida de dispersión muy tosca y puede ser engañosa si hay una puntuación atípicamente alta o baja. El rango falla también al dar indicaciones sobre la variabilidad de las puntuaciones en torno a la media de la distribución. Algunas veces los investigadores usarán el *rango intercuartílico*, que indica la dispersión entre la mitad intermedia de las puntuaciones.

Desviación típica

La **desviación típica** es un índice numérico que indica la variabilidad promedio de las puntuaciones. Nos habla, en otras palabras, sobre la distancia, sobre el promedio, de las puntuaciones desde la media. Una distribución que tiene un conjunto de puntuaciones relativamente heterogéneo que se separa ampliamente de la media (por ejemplo, la clase A de la figura 5.7) tendrá una desviación típi-

ca más grande que un conjunto homogéneo de puntuaciones que se agrupan alrededor de la media (Clase B de la figura 5.7). El primer paso para calcular la desviación típica (abreviada DT, σ [sigma], o s) es encontrar la distancia entre cada puntuación y la media (ver figura 5.8), de esta forma determinamos la cantidad en que cada puntuación se desvía o difiere de la media. En un sentido, la desviación típica simplemente es el promedio de todas las puntuaciones de desviación, la distancia promedio de las puntuaciones desde la media.

Luego, para cualquier conjunto de puntuaciones, puede calcularse la desviación típica, que será única para la distribución e indica la cantidad, sobre el promedio, que el conjunto de puntuaciones se desvía de la media. (El Apéndice C repasa los pasos para calcular la desviación típica. Los pasos no son complicados). El acuerdo más común al presentar la desviación típica es indicar que una desviación típica es igual a algún número (por ejemplo, $DT = 15.0$; $s = 3.40$). Una desviación típica añadida o suprimida a partir de la media tiene un significado especial; nos habla acerca de la distancia a la que están de la media la mayoría, pero no todas, las puntuaciones. Por ejemplo, el 68% de las puntuaciones estará dentro de la primera desviación típica en una distribución normal. Esta propiedad de la desviación típica se ilustra en la figura 5.9, en que $1 DT = 5$. Observe que a ambos lados de la media (15) hay una línea que designa $-1 DT$ y $+1 DT$. Las direcciones positiva y negativa desde la media son equivalentes en unidades de puntuación (ambas $-$ y $+1 DT = 5$ unidades) y entre -1 y $+1 DT$ hay alrededor del 68% del número total de las puntuaciones en la distribución. Si asumimos que la distribución es normal, entonces el 50% de las puntuaciones está por encima de la media y el 50% por debajo de ella. Ahora, ya que sabemos que hay un número igual de puntuaciones en cada lado de la media, sabemos que el 34% de las puntuaciones debería estar entre la media y $-$ o $+1 DT$, y si el 50% de las puntuaciones está por debajo de la media y añadimos el 34% por ascenso de $+1 DT$, entonces sabemos que alrededor del 84% de las puntuaciones de la distribución está por debajo de $+1 DT$. De forma similar, si sustraemos 34 de 50, sabemos que el 16% de las puntuaciones está por debajo de $-1 DT$.

FIGURA 5.8: ILUSTRACIÓN DE LA DISTANCIA DE CADA PUNTUACIÓN DESDE LA MEDIA

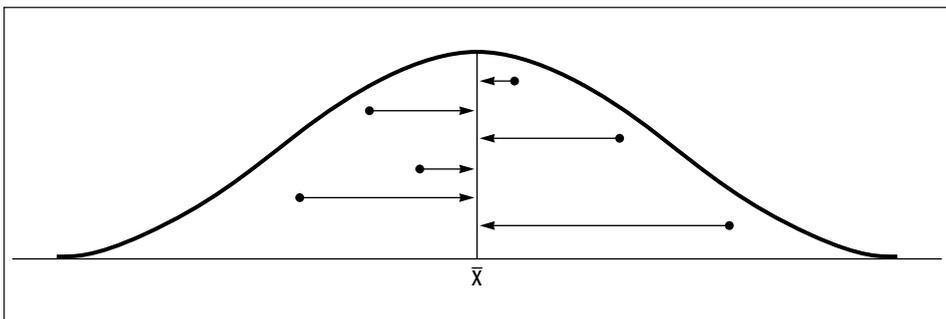
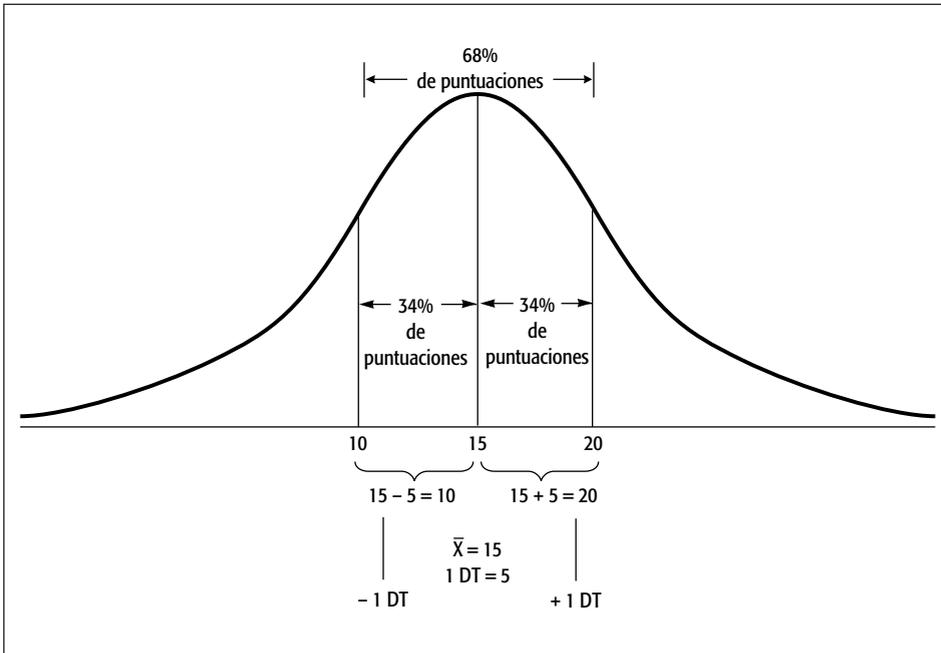


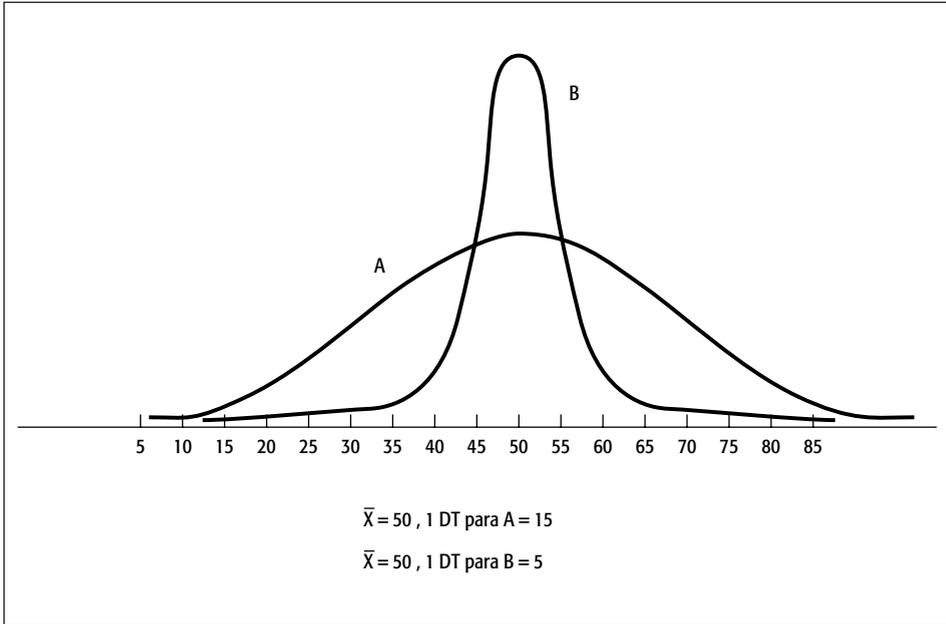
FIGURA 5.9: IRELACIÓN DE LA DESVIACIÓN TÍPICA CON EL RANGO PERCENTIL EN UNA DISTRIBUCIÓN NORMAL



Cuando indicamos que un cierto porcentaje de las puntuaciones está en o por debajo de una puntuación particular, nos referimos al **rango percentil** de la puntuación. Si, por ejemplo, una puntuación de 38 está en el percentil 87, significa que el 87% de las puntuaciones son iguales o menores que 38. En otras palabras, sólo el 12% de las puntuaciones son más altas que 38. Con distribuciones normales $+1 \text{ DT}$ siempre está en el percentil 84 y -1 DT está en el percentil 16.

La interpretación de 1 DT siempre es la misma con respecto al porcentaje de puntuaciones dentro de ciertos puntos de una distribución normal. Sin embargo, debido a que las unidades numéricas que se usan para representar puntuaciones cambian, la desviación típica puede equivaler a 15 en una distribución y .32 en otra distribución. O, en una circunstancia con las mismas unidades numéricas pero dos distribuciones diferentes, las desviaciones típicas serán únicas para cada distribución. Esto es, una DT tiene un significado que es constante para cualquier distribución independientemente del valor real de 1 DT para cada distribución. Por ejemplo, en la figura 5.10 se ilustran dos distribuciones. La distribución A tiene una desviación típica grande, la B una pequeña; una puntuación de 65 en la distribución A tiene el mismo rango de percentil que 55 en la distribución B.

FIGURA 5.10: COMPARACIÓN DE DISTRIBUCIONES CON DESVIACIONES TÍPICAS DIFERENTES



Además de la media, la desviación típica es una forma excelente para indicar la naturaleza de la distribución en un conjunto de puntuaciones. En investigación, habitualmente, se presenta la desviación típica junto con la media. Una medida de dispersión relacionada con la desviación típica se denomina **varianza** de una distribución (simbolizada como σ^2 o s^2 ; por esto, la desviación típica es igual a la raíz cuadrada de la varianza). Sin embargo, el término *varianza* se emplea frecuentemente como un término general en relación con la dispersión (por ejemplo, definiendo que la varianza es grande o pequeña) y raramente se presenta como un número específico para indicar variabilidad.

En los extractos 5.5 y 5.6 hay ejemplos de la forma en que puede presentarse una desviación típica. Las desviaciones típicas casi siempre se presentan junto a las medias. En el extracto 5.5, para cada uno de los factores que se usan para calificar a los alumnos, hay una media y una desviación típica para los profesores de secundaria, una media y una desviación típica para los profesores de instituto y una media y una desviación típica para la muestra total. Observe que los valores de la desviación típica varían de .76 a 1.32 mostrando diferentes grados de varianza. Las medias al final de la escala están asociadas a desviaciones típicas más pequeñas. En las escalas Likert, como por ejemplo ésta, son normales las desviaciones típicas cercanas a 1. En el extracto 5.6, las desviaciones típicas están entre paréntesis.

EXTRACTO 5.5 DESVIACIÓN TÍPICA

TABLA 12: Medias y desviaciones típicas de los ítems de factores que se usan para calificar las prácticas por profesores de secundaria¹

Factores	Medio (N 5 630)		Alto (N 5 846)		Total (N 5 1476)	
	Media	DT	Media	DT	Media	DT
Actuación disruptiva del alumno	1.5	.83	1.60	.91	1.56	.88
Mejorar desde el comienzo del año	2.86	1.14	2.83	1.12	2.85	1.13
Esfuerzo del estudiante –cuánto ha intentado aprender	3.31	1.13	3.16	1.10	3.23	1.11
Niveles de capacidad de los alumnos	3.38	1.33	3.43	1.28	3.41	1.30
Hábitos de trabajo y limpieza	2.80	1.07	2.68	1.06	2.73	1.07
Finalización de deberes de casa (sin calificar)	3.02	1.06	2.95	1.12	2.98	1.10
Calidad de los deberes completados (calificada)	3.18	1.15	3.22	1.14	3.20	1.15
Rendimiento académico como opuesto a otros factores	4.37	1.08	4.34	1.09	4.35	1.08
Rendimiento comparado con otros alumnos de la clase	2.06	1.13	2.23	1.18	2.16	1.17
Rendimiento comparado con una escala de porcentaje correcto	4.44	1.24	4.45	1.31	4.43	1.29
Objetivos de aprendizaje específico adquiridos	4.38	.92	4.35	.91	4.37	.92
Grado en el que los alumnos prestan atención y/o participan en clase	3.12	1.11	3.20	1.12	3.17	1.12
Inclusión de ceros determinando el porcentaje correcto final.	3.61	1.29	3.90	1.32	3.77	1.12
Crédito extra por rendimiento no académico	1.54	.86	1.49	.76	1.51	.80
Crédito extra por rendimiento académico	2.66	1.18	2.54	1.06	2.60	1.11

¹Se usó una escala de seis puntos en la que 1 = no en absoluto y 6 = completamente

Fuente: McMillan, J. H. y Workman, D. (1998), *Teachers' Classroom assessment and Grading Practices*, Richarmon, VA: Metropolitan Educational Research Consortium y Virginia Commonwealth University.

Diagrama de caja (*box-and-whisker plot*)

El **diagrama de caja** se usa para dar una imagen de la variabilidad. Se forma una «caja» por cada variable. El tamaño de esta caja rectangular se determina por el primer y tercer cuartil de la distribución (del percentil 25 al 75). Los «bigotes» son líneas que se dibujan desde los extremos del rectángulo hasta los percentiles 10 y 90. Algunas veces se incluyen puntos adicionales para mostrar puntuaciones extremas altas o bajas. El diagrama de caja del extracto 5.7 muestra cómo se compara el rendimiento de los estudiantes norteamericanos en matemáticas con el rendimiento de los estudiantes japoneses. Preste atención a la gran variación del rendimiento entre los estudiantes norteamericanos.

EXTRACTO 5.6 DESVIACIÓN TÍPICA

Medias y desviaciones típicas para todas las variables

VARIABLES DEPENDIENTES

Cualquier carrera de ciencias	4.40	(1.55)
Profesionales de ciencias físicas ¹	4.24	(1.79)
Profesionales de ciencias de la salud	4.56	(2.01)
Profesionales de servicios humanos	4.22	(1.77)

VARIABLES INDEPENDIENTES

Ciencia GPA ²	3.66	(.48)
Apoyo de los amigos	5.69	(1.08)
Número de actividades de ciencias/matemáticas ³	1.73	(1.32)
Número de actividades que no son de ciencias ⁴	9.74	(3.67)
Percepciones de las madres sobre la capacidad en ciencias del niño	6.10	(1.05)
Valoración de las madres de la ciencia para las mujeres	4.57	(1.04)
Interés de los adolescentes en biología	5.10	(1.43)
Interés de los adolescentes en ciencias físicas	4.75	(1.40)

¹ El profesional incluye trabajos que precisan licenciaturas y grados avanzados. ²Rango = 1-4.

³ Rango = 0-5. ⁴Rango = 2-21.

Fuente: Jacobs, J. E., Finken, L. L., Griffin, N. L., y Wright, J. D. (1998), «The career plans of science-talented rural adolescents girls», *American Educational Research Journal*, 35(4), 681-704.

Puntuaciones típicas

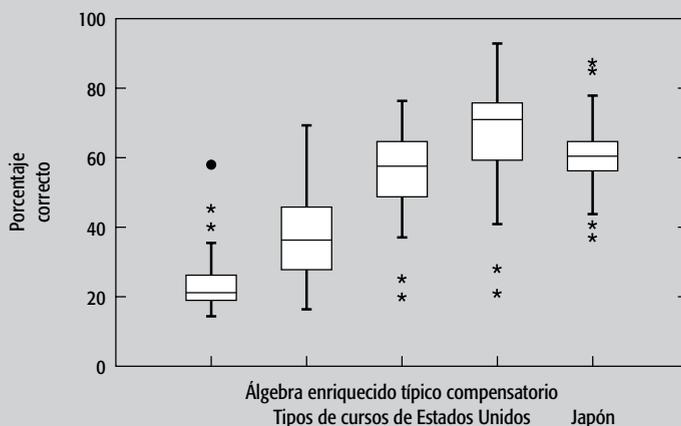
Puede observar que resulta engorroso analizar varias distribuciones si las medias y las desviaciones típicas son diferentes para cada distribución. Para atenuar este problema y facilitar la interpretación, las distribuciones de puntuaciones

directas se convierten en puntuaciones típicas. Las **puntuaciones típicas** tienen un significado constante normativo y relativo. Son puntuaciones obtenidas a partir de la media y la desviación típica de la distribución de puntuaciones directas.

Debido a que, como hemos visto, una distribución normal tiene ciertas propiedades que son útiles para comparar la puntuación de una persona con las de otras, en la conversión a puntuaciones típicas normalizadas se pueden asumir las propiedades de la curva normal. De esta manera, las distribuciones de puntuaciones directas con diferentes medias y desviaciones típicas, que son difíciles de comparar, pueden transformarse en las mismas puntuaciones típicas y compararse fácilmente. Dado que las puntuaciones típicas son transformaciones lineales, es concebible que una pequeña diferencia de puntuación directa se exagere cuando se convierte en puntuación típica. Por ejemplo, el *Scholastic Assessment Test*¹ tiene una puntuación típica media de 500 y una desviación típica de 100, mientras que las puntuaciones directas son mucho menores. Por esto, una diferencia de puntuación directa de 2 o 3 preguntas puede resultar en una diferencia de puntuación típica de 10 o 20 «puntos».

EXTRACTO 5.7
DIAGRAMA DE CAJA

Dadas las diferencias de los patrones de dominio del álgebra entre estos tipos de curso en Estados Unidos, ¿qué ocurre cuando el rendimiento de los estudiantes estadounidenses en álgebra se analiza separadamente según el tipo de curso? La figura 3 representa tal disgregación por las puntuaciones de un postest nivel-clase (ampliado para incluir el postest paralelo de rendimiento para Japón) y muestra un patrón específico.



(continúa)

¹ (Nota del revisor: examen de evaluación académica, también conocido como SAT. Constituye el estándar de las pruebas de admisión que suelen solicitar los colleges y universidades en EEUU).

(continuación)

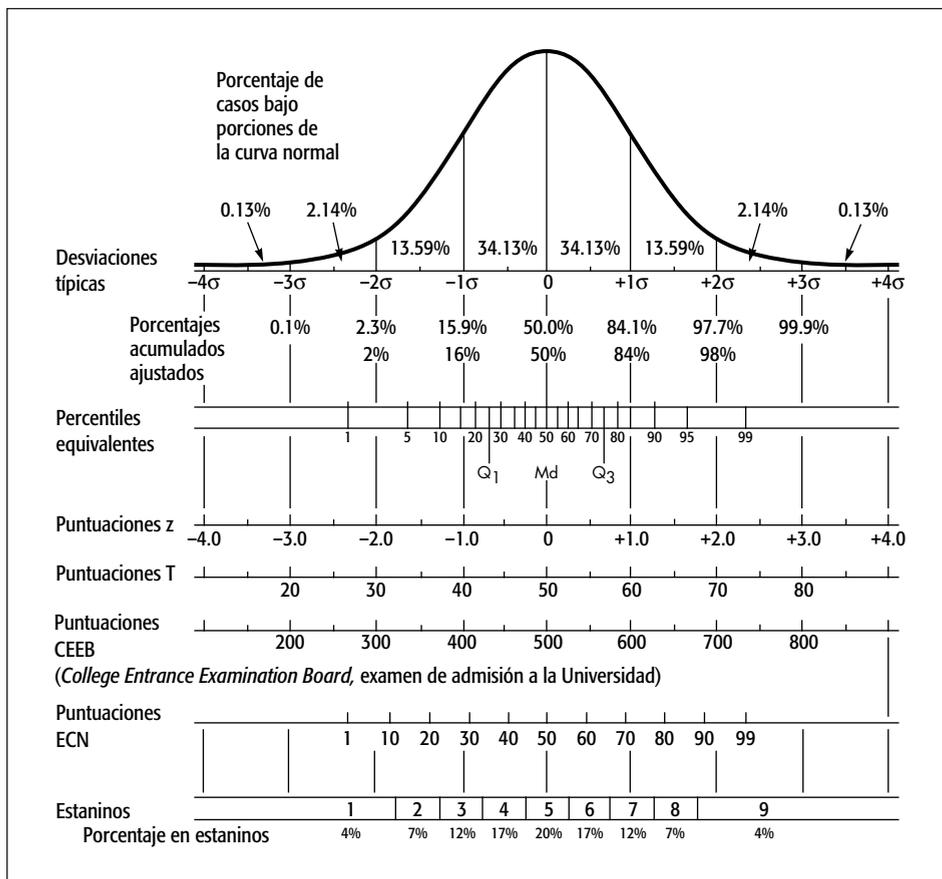
FIGURA 3. Estados Unidos y Japón: postest de rendimiento en álgebra entre la población A (Nota: en las cajas como las que encontramos en las figuras 3, 4 y 5, la longitud de la caja, el rectángulo limitado por las «bisagras», representa la proporción de la distribución que cae entre los percentiles 25 y 75. La línea que atraviesa la caja representa la mediana. La longitud de los «bigotes» representa el mínimo y el máximo o el valor exterior adyacente,

$$1.5 = (\text{percentil}_{75} - \text{percentil}_{25}),$$

Si esto es menos que el mínimo y el máximo. El * y el • representan los valores extremos.

Fuente: Westbury, I. (1992), «Comparing American and Japanese achievement: Is the United States really a low achievement?», *Educational Researcher*, 2 (15), 18-24.

FIGURA 5.11: CURVA NORMAL, DESVIACIONES TÍPICAS, PERCENTILES Y PUNTUACIONES TÍPICAS SELECCIONADAS



Fuente: Seashore, Harold G. (1980), «Methods of expressing test scores», en *Test Service Notebook 148*. New York: The Psychological Corporation.

La **puntuación z** es la puntuación típica más básica, con una media de 0 y una desviación típica de 1. De esta forma, una puntuación z de +1 está en el percentil 84 para una distribución normal, -1 en el percentil 16 y -2 se encuentra en el percentil 2. Otras puntuaciones típicas son transformaciones lineales de la puntuación z, con medias y desviaciones típicas seleccionadas arbitrariamente. Es decir, es posible elegir cualquier media y cualquier desviación típica. Por ejemplo, la mayoría de los tests de CI usan 100 como media y de 15 a 16 como desviación típica. La puntuación CI resultante es una puntuación típica (el coeficiente CI, edad mental dividida entre edad cronológica $\times 100$, hoy en día es raro que se use). La figura 5.11 muestra una distribución normal, la desviación típica, percentiles y algunas puntuaciones típicas comunes.

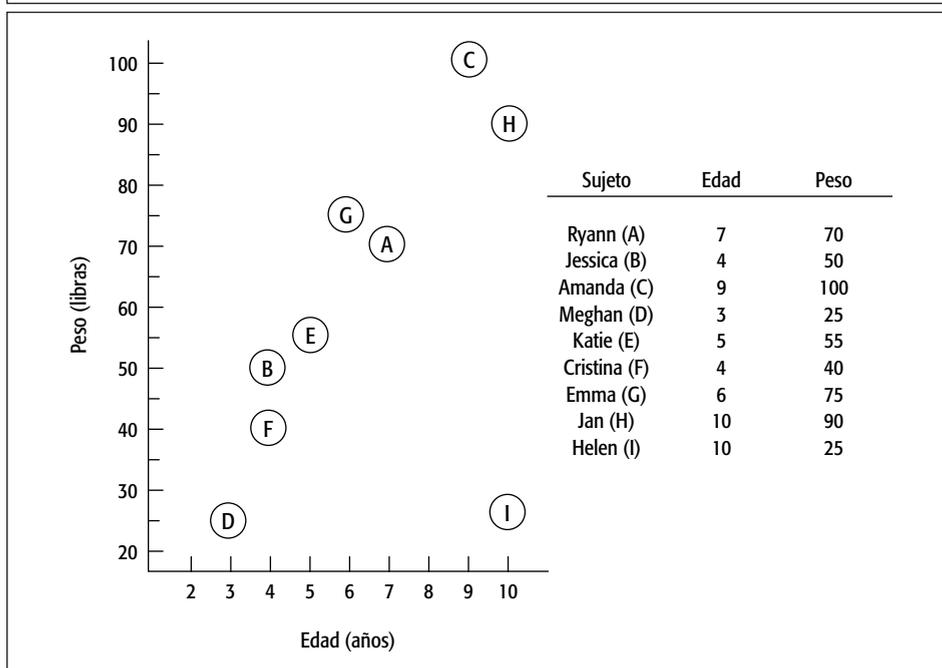
MEDIDAS DE RELACIÓN

Hasta este punto hemos discutido la estadística descriptiva que se utiliza para resumir o dar una imagen de grupos sobre una variable cada vez. Sin embargo, existen muchas cuestiones de interés que dependen de la forma en que dos o más variables se relacionan entre sí. ¿Están más motivados los alumnos más brillantes? Si incrementamos la frecuencia del refuerzo, ¿aumentará también el comportamiento diana o reforzado? ¿Existe una relación entre la autoestima y el rendimiento? Si los alumnos aumentan su esfuerzo al estudiar, ¿se sentirán más satisfechos de su rendimiento? En cada caso, se miden dos variables para cada sujeto del grupo.

Diagrama de dispersión

La medida más fundamental de relación se llama *diagrama de dispersión*. El **diagrama de dispersión** es un gráfico de representación de la relación, que se logra al presentar visualmente la intersección de las puntuaciones de cada sujeto en las dos variables. Como se muestra en la figura 5.12, una variable se ordena en el eje horizontal (por ejemplo, edad) y la segunda variable se ordena en el eje vertical (peso). Las puntuaciones de cada sujeto se indican al lado del gráfico en orden aleatorio y las intersecciones se indican con la letra asignada a cada sujeto. Todas las intersecciones juntas, forman un patrón que proporciona una indicación general de la naturaleza de la relación. Obviamente, cuando los niños crecen, sus pesos se incrementan y en estos casos la relación se denomina positiva o directa. De este modo, en una **relación positiva** el aumento del valor de una variable va acompañado del incremento del valor en la segunda variable. Recíprocamente, cuando el valor de una variable disminuye, el valor de la otra variable también lo hace.

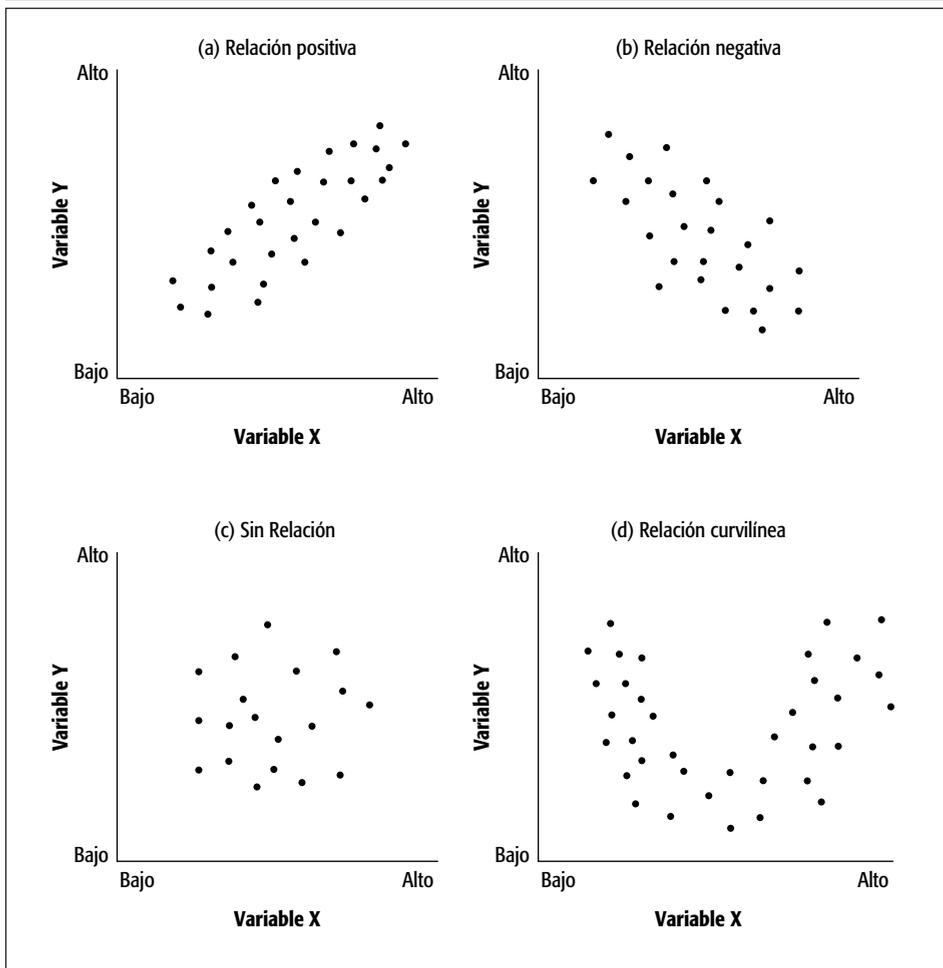
FIGURA 5.12: DIAGRAMA DE DISPERSIÓN



Los diagramas de dispersión son útiles para la identificación de puntuaciones que son muy atípicas (puntuaciones extremas, también conocidas como *outliers*) comparadas con el patrón global. Por ejemplo, en la figura 5.12, Hellen tiene 10 años y un peso de 25 libras, lo que es bastante diferente de lo que se presenta entre los puntos A y H. En tales casos, el investigador podría buscar errores en la puntuación, medida o registro de los datos, debido a que las puntuaciones extremas son insólitas. Los diagramas de dispersión también proporcionan un primer indicio de si la relación es lineal o curvilínea (ver figura 5.13). (La propuesta más común en las relaciones gráficas es usar puntos, no círculos, en las intersecciones dentro del gráfico.)

En los diagramas de dispersión pueden aparecer varios tipos diferentes de patrones. Cuando una variable disminuye y la otra aumenta (por ejemplo, la cantidad de kilómetros de un neumático y la profundidad de la huella que le queda), existe una **relación negativa** o *inversa*. Si no hay un patrón en todo el gráfico, entonces no existe relación. La figura 5.13 muestra diferentes diagramas de dispersión. Observe la relación curvilínea en la figura 5.13 (d). Las relaciones curvilíneas no son infrecuentes, pero normalmente sólo se detectan dibujando los puntos. Un ejemplo de una relación curvilínea puede ser el nivel de ansiedad durante la realización de un test. El comportamiento podría ser a menudo bajo durante niveles de ansiedad altos como bajos y altos para niveles medios de ansiedad.

FIGURA 5.13: DIAGRAMAS DE DISPERSIÓN DE RELACIONES

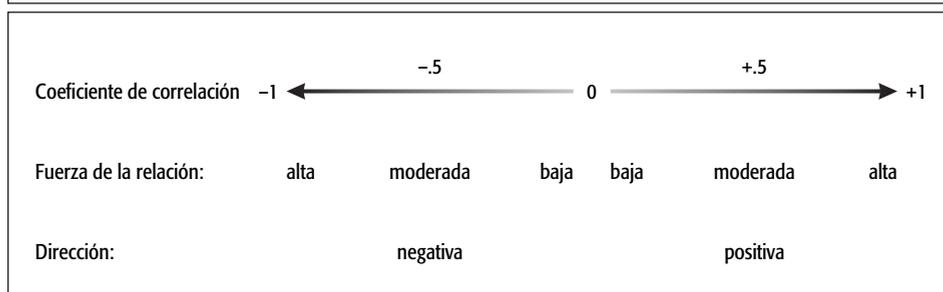


Por lo tanto, la dirección del patrón en el diagrama de dispersión indica si existe una relación y si ésta es positiva, negativa o curvilínea. Si dibujamos una línea a través de los puntos para minimizar la distancia de cada punto a la línea, entonces, el grado de agrupación alrededor de la línea indica la fuerza de la relación. Los puntos que están más dispersos tienen una relación débil, mientras que los puntos agrupados cerca de la línea indican una relación alta o fuerte. La fuerza de la relación es independiente de su dirección. Puntos agrupados tan estrechamente que forman una línea recta representan una relación perfecta (máxima fuerza). De este modo, las correlaciones indican tres cosas: si existe una relación, la dirección de la relación y la fuerza de la relación.

Coeficiente de correlación²

Aun cuando los diagramas de dispersión son herramientas indispensables para la evaluación de la relación entre dos variables, los investigadores muy rara vez presentan este tipo de gráficos en los artículos publicados. El acuerdo más común es calcular un número para representar la relación, denominado **coeficiente de correlación**. Hay muchos tipos de coeficientes de correlación y la elección de uno u otro se determina por la escala usada en la recogida de datos y la pregunta de investigación. Sin embargo, la interpretación del número elegido, básicamente es la misma. El número que representa la correlación puede ir de -1.00 a $+1.00$. Un valor positivo alto (por ejemplo, $.85$, $.90$, $.96$) representa una relación positiva alta; un valor positivo bajo ($.15$, $.20$, $.08$) una relación positiva baja; un valor negativo moderado (por ejemplo, -0.40 , -0.37 , -0.52) una relación negativa moderada, un valor de 0 que no hay relación y así sucesivamente. De esta manera, la fuerza de la relación es mayor cuando la correlación se aproxima a $+1$ o -1 desde 0 . Esto se ilustra en la figura 5.14. Observe que la fuerza de la relación es independiente de la dirección.

FIGURA 5.14: RELACIÓN DE FUERZA Y DIRECCIÓN DE LAS CORRELACIONES



La correlación técnica más común de es el coeficiente de correlación de producto-momento de Pearson (representado por r), y la correlación se indica por $r = .65$, $r = .78$, $r = .03$, etc. (Observe que no existe un signo más delante de los valores positivos, debido a que hay un signo negativo delante de los valores negativos). El cociente de correlación de producto-momento se utiliza cuando ambas variables usan escalas continuas, como por ejemplo, puntuaciones de tests de rendimiento, la nota media, los inventarios de autoestima y la edad. Ya que las

² Esta presentación se limita a la correlación simple. Los procedimientos correlacionales más avanzados, como por ejemplo correlación múltiple, correlación parcial, análisis de función discriminante y correlación canónica, se basan en estos principios para examinar las relaciones combinadas de diversas variables.

puntuaciones pueden ser presentadas también de forma dicotómica, en distintas categorías o rangos, se usan para medir la relación otras técnicas de correlación dependiendo de la escala empleada para medir las variables. Algunas de estas técnicas se resumen en la tabla 5.5.

Los extractos 5.8 y 5.9 muestran cómo presentar datos correlacionales. En el extracto 5.7, aparece una lista de todas las correlaciones de interés en el estudio. En el extracto 5.8 hay una **matriz de correlaciones**, en la que muchas variables se correlacionan entre sí. Los números en una fila al principio de la tabla se corresponden con las variables que están listadas verticalmente a la izquierda. La correlación entre el rechazo de los profesores y las críticas hacia su motivación es $-.54$; la correlación del compromiso con las expectativas a largo plazo de los profesores es $.44$.

En el capítulo 7 comentaremos los principales principios de la interpretación de los coeficientes de correlación.

TABLA 5.5: TIPOS DE COEFICIENTES DE CORRELACIÓN

Tipo de coeficiente	Símbolo	Tipos de variables
Producto-momento Pearson	r	Ambas continuas
Spearman	r_s	Ambas ordinales
Biserial	r_b	Una continua, otra dicotomizada artificialmente
Biserial puntual	r_{pb}	Una continua, otra dicotómica
Tetacórica	r^t	Ambas dicotomizadas artificialmente
Coefficiente Fi	ϕ	Ambas dicotómicas
Coefficientes de contingencia	C	Ambas con 2 o más categorías
Razón de correlación, eta	η	Ambas continuas (usadas en relaciones curvilíneas)

EXTRACTO 5.8

CORRELACIÓN PRODUCTO-MOMENTO DE PEARSON

Se calcularon correlaciones producto-momento de Pearson entre todas las variables demográficas (por ejemplo, porcentaje de blancos, bajos ingresos, tasa de asistencia, porcentaje de movilidad, tasa de abandono de alumnos de instituto, graduación en el instituto, tamaño medio de la clase, años de experiencia de los profesores, ratio alumno-profesor, salario medio del profesor, gasto medio por alumno) y las puntuaciones

(continúa)

(continuación)

de rendimiento (por ejemplo, puntuaciones en lectura y matemáticas). Las correlaciones de 1994 (ver tabla 2) fueron similares en los cursos y asignaturas con los datos de 1994.

Para resumir, detectamos relaciones estadísticamente significativas (o asociaciones) entre las variables demográficas escolares y las puntuaciones de rendimiento. Las relaciones más fuertes afectaban a las siguientes variables: ingresos bajos, porcentaje de blancos, graduación en el instituto y abandono de alumnos de instituto. Existían relaciones moderadas para la asistencia, movilidad y tasa profesor-alumno en el instituto. Las relaciones más débiles afectaban al tamaño medio de la clase, la ratio profesor-alumno en primaria, el salario del profesor, la experiencia del profesor y el gasto por alumno.

TABLA 2: Correlaciones de medidas de actitud con puntuaciones de rendimiento

Variable	3 ^{er} curso		10 ^o curso	
	Lectura	Matemáticas	Lectura	Matemáticas
Porcentaje de blancos	.78	.66	.75	.67
Ingresos bajos	-.79	-.72	-.79	-.75
Asistencia	.59	.53	.82	.72
Movilidad	-.52	-.46	-.54	-.49
Abandono	—	—	-.69	-.61
Graduación Instituto	—	—	.76	.69
Tamaño medio de clase, 3 curso	-.09**	-.06**	—	—
Tamaño medio de clase, Instituto	—	—	-.18	-.11**
Experiencia del profesor	-.14	-.13	-.05	.00
Tasa profesor-alumno primaria	-.26	-.22	-.32	-.24
Tasa profesor-alumno instituto	—	—	-.32	-.24
Salario del profesor	-.20	-.08	-.05	.07
Gasto por alumno	-.31	-.19	.10*	-.01

Nota: 3^{er} curso, n = 2,307; 10^o curso, n = 644. Las correlaciones son estadísticamente significativas en el nivel .001 a menos que se señale otra cosa.

*p < .05. **p < .01.

Fuente: Sutton, A., y Soderstrom, I. (1999), «Predicting elementary and secondary school achievement with school-related and demographic factors», *Journal of Educational Research*, 92 (6), 330-338.

EXTRACTO 5.9
MATRIZ DE CORRELACIÓN

En la tabla 2 se describen las correlaciones de orden cero entre las variables de contexto motivador y los signos de alienación. Todas las correlaciones estaban en la dirección esperada. Los problemas disciplinarios de los alumnos estaban fuertemente relacionados con sus informes sobre el desinterés y la crítica y expectativas de los profesores. La relación más fuerte encontrada ocurría entre las percepciones de los alumnos de las expectativas de los profesores y el compromiso de los alumnos. Las aspiraciones académicas de los compañeros y sus percepciones de las limitaciones económicas de la educación estaban relacionadas tanto con los problemas de disciplina como con el compromiso.

TABLA 2: Correlaciones de orden-cero entre las variables de contexto motivador y los índices de alineación

Variables de contexto motivador	1	2	3	4	5	6	7	8	9	10
1. Desinterés y crítica de los profesores	—									
2. Motivación de los profesores	-.54***	—								
3. Expectativas a largo plazo de los profesores	-.39***	.34***	—							
4. Aspiraciones académicas del grupo	-.21***	.13**	.36***	—						
5. Resistencia a las normas del colegio de los compañeros	.27***	-.09	-.22***	-.47***	—					
6. Apoyo académico del grupo	-.32***	.31***	.29***	.47***	-.44***	—				
7. Limitaciones económicas de la educación	.38***	-.24***	-.35***	-.38***	.27***	-.25***	—			
8. Beneficios económicos de la educación	-.12*	.21***	.32***	.32***	-.12*	.21***	-.36***	—		
Índices de alienación										
9. Problemas de disciplina	-.35***	.15**	-.36***	-.26***	.17**	-.21***	.29***	-.11**	—	
10. Compromiso	-.16**	.04	.44***	.27***	-.14**	.19***	-.22***	.11**	-.47***	—

*p < .05. **p < .01. ***p < .001.

Fuente: Murdock, T. B. (1999), «The social context of risk: Status and motivational predictors of alienation in middle school», *Journal of Educational Psychology*, 91 (1), 62-75.

RESUMEN

Este capítulo ha presentado los principios fundamentales de la estadística descriptiva. Los procedimientos estadísticos se usan de una u otra forma en casi todos los estudios de investigación cuantitativos. Los principales puntos que se han tratado se resumen a continuación:

1. La estadística descriptiva está constituida por índices que resumen o caracterizan un gran número de observaciones.
2. Las escalas de medida (nominal, ordinal, intervalo y de razón) y el propósito de la investigación sugieren la estadística descriptiva apropiada.
3. Las distribuciones de frecuencia en forma de intervalos de clase, histogramas, polígonos de frecuencia y gráficos de tallo y hojas proporcionan una imagen global de todos los datos.
4. Las medidas de tendencia central incluyen la media, la mediana y la moda. Cada medida proporciona un índice numérico de la puntuación típica en la distribución.
5. La media es la mejor medida de tendencia central para distribuciones que no tienen puntuaciones extremadamente altas o bajas; la mediana es la mejor para datos altamente asimétricos. A menudo, suelen presentarse ambas, la media y la mediana.
6. Las medidas de variabilidad indican la dispersión de las puntuaciones desde la media de la distribución.
7. Las medidas de variabilidad incluyen el rango, la desviación típica y los diagramas de caja.
8. La desviación típica es una medida de variabilidad exclusiva de cada distribución que indica, sobre el promedio, cuánto se desvían las puntuaciones desde la media.
9. Las puntuaciones típicas son distribuciones de puntuación directa convertidas en unidades comunes para indicar la media y la desviación típica.
10. Los gráficos de dispersión se utilizan para indicar la dirección general y la fuerza de una relación entre dos variables de un grupo o muestra.
11. Los coeficientes de correlación son números que representan la dirección y fuerza de la relación entre dos o más variables.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. ¿Cuál de los siguientes enunciados es verdadero (más de uno puede ser verdadero) con relación a la estadística descriptiva?

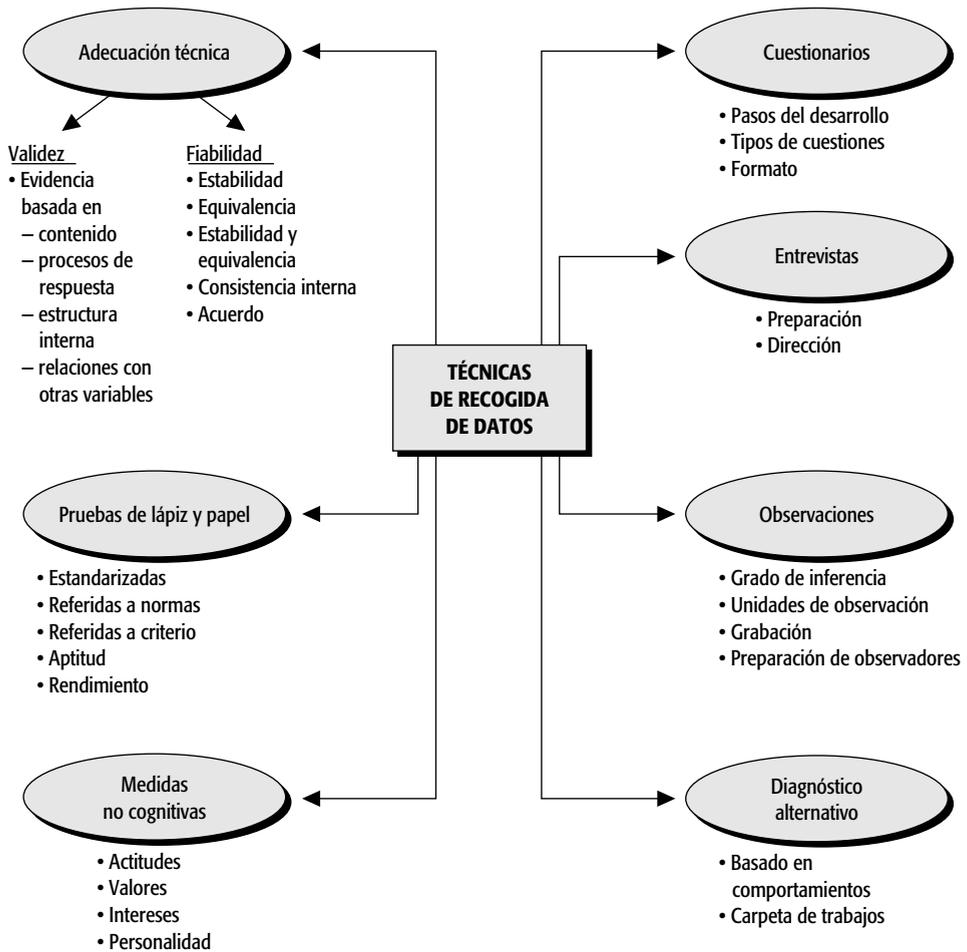
- a. La estadística descriptiva se usa para resumir los datos.
 - b. La estadística descriptiva se emplea para inferir características de la población a partir de una muestra.
 - c. La estadística descriptiva no es necesaria para interpretar la estadística inferencial.
 - d. Los coeficientes de correlación son estadística descriptiva.
 - e. La elección de la estadística descriptiva depende del propósito de la investigación y el tamaño de la muestra.
2. Las escalas de medida que proporcionan categorías sin orden se llaman...
 - a. intervalo.
 - b. no ordinales.
 - c. nominal.
 - d. razón.
 3. Los percentiles son ejemplos de una escala de medida llamada...
 - a. nominal.
 - b. intervalo.
 - c. ordinal.
 - d. razón.
 4. Las frecuencias de distribución se representan por todos los siguientes procedimientos excepto por...
 - a. histogramas.
 - b. polígonos.
 - c. curva normal.
 - d. agrupamiento por rangos
 5. La _____ es la medida de tendencia central más ampliamente usada.
 - a. media
 - b. mediana
 - c. desviación típica
 - d. moda
 6. Dadas las siguientes puntuaciones: 4, 7, 7, 8, 10, 12, 12, 12, 18; la media es _____; la mediana es: _____; y la moda es _____.
 - a. 8, 12, 10
 - b. 8, 10, 12
 - c. 10, 10, 12
 - d. 10, 10, 10
 - e. 12, 12, 12
 7. La cantidad promedio de dispersión de un conjunto de puntuaciones alrededor de la media se evalúa numéricamente por...
 - a. el rango.
 - b. la desviación típica.
 - c. los rangos de percentil.

- d. la mediana.
8. Más desviación típica está siempre aproximadamente en el percentil _____ en una distribución normal.
- 60
 - 50
 - 90
 - 84
9. ¿Cuál de las siguientes correlaciones muestra la relación más fuerte?
- .50
 - .75
 - .67
 - .30
10. Una buena razón para presentar gráficamente correlaciones es encontrar...
- relaciones curvilíneas.
 - relaciones positivas.
 - relaciones negativas.
 - la a y la b.

Problemas aplicados

- Para cada caso, elegir el procedimiento estadístico más apropiado.
 - Un profesor de un grupo con nivel bajo de lectura quiere saber cuál es la puntuación promedio para el grupo de 25 alumnos.
 - Una administración quiere descubrir si hay relación entre el absentismo de los profesores y el rendimiento de los alumnos.
 - Un profesor de matemáticas quiere saber cuántos grupos de capacidad debería formar en una clase de 30 alumnos.
 - Un profesor está interesado en el número de alumnos que valoran su actuación como buena, excelente, media o baja.
- Identificar la escala de medida en cada una de las siguientes:
 - Actitudes hacia el colegio.
 - Agrupación de los alumnos sobre la base de su color de pelo.
 - Pidiendo a jueces que clasifiquen a los alumnos desde los más cooperativos hasta los menos cooperativos.

Técnicas de recogida de datos cuantitativos



PALABRAS CLAVE

constructo infraestimado
varianza residual del constructo
evidencia basada en el contenido del test
evidencia basada en los procesos de respuesta
evidencia basada en la estructura interna
evidencia basada en las relaciones con otras variables
fiabilidad
estabilidad
equivalencia
consistencia interna por dos mitades
Kuder-Richardson
Alpha de Cronbach
Acuerdo
tests estandarizados referido a normas
referido a criterio
test de aptitud
pruebas de rendimiento diagnóstico alternativo
diagnóstico basado en comportamientos
carpeta de trabajos no cognitivo
deseabilidad social
cuestionario
preguntas con dos objetivos
formato cerrado
formato abierto
escala
escala de Likert
diferencial semántico
lista de comprobación
preguntas de contingencia
preguntas estructuradas
preguntas semiestructuradas
preguntas no estructuradas
pregunta destacada
análisis
observador completo
inferencia alta
inferencia baja
duración grabación
intervalo de grabación
observación continua
muestreo de tiempo
no reactiva
medidas discretas

Este capítulo detalla los principios de validez y fiabilidad como consideraciones importantes de la calidad de las medidas y presenta las características de las cinco principales técnicas para la recogida de datos cuantitativos: tests, cuestionarios, entrevistas, observación y medidas discretas. Se debaten las ventajas y desventajas de cada tipo con relación a los objetivos de los problemas de investigación y de las pruebas que se utilizan con cada técnica. Estas técnicas se utilizan en diferentes tipos de investigación, como se define tanto por el propósito

como por el diseño de investigación. Una vez que el propósito y las limitaciones de la investigación están claros, se elige una técnica determinada que se ajuste al diseño de investigación. No siempre una técnica sencilla es la mejor, la más fácil o la más conveniente.

FUNDAMENTOS DE MEDICIÓN CUANTITATIVA: ADECUACIÓN TÉCNICA

La medición cuantitativa precisa algún tipo de procedimiento o dispositivo para obtener índices numéricos que se corresponden con las características de los sujetos. Luego, los valores numéricos se resumen y se presentan como los resultados del estudio. Consecuentemente, los resultados dependen en gran medida de la calidad de la medición. Si la medida es débil o sesgada, entonces así serán los resultados. Por el contrario, medidas fuertes, incrementan la confianza de que los hallazgos son exactos. Por ello, es imperativo, entender lo qué hace una medición «fuerte» o «débil». Si necesita elegir pruebas para dirigir un estudio o para evaluar los resultados, es necesario entender cómo afectan a la calidad de la medida. En esta sección se argumentan dos conceptos técnicos de medición, validez y fiabilidad, como criterios importantes a la hora de determinar la calidad.

Validez

Como se indicó en el capítulo 4, la validez de un test es el grado en qué las deducciones realizadas sobre la base de puntuaciones numéricas resultan apropiadas, significativas y útiles. Validez es un juicio sobre la idoneidad de una medida para las conclusiones o decisiones específicas que son resultado de las puntuaciones generadas. En otras palabras, la validez es un concepto específico de situación: la validez se evalúa dependiendo de la finalidad, de la población y de las características ambientales en las que se realiza la medición. Por lo tanto, el resultado de un test puede resultar válido en una situación e inválido en otra. Consecuentemente, para asegurar a los demás que el proceso tiene validez con relación a los problemas de investigación, los sujetos y el entorno del estudio, es procedimientos que se utilizaron para recoger los datos.

Esta conceptualización de la validez del test implica mucho más que determinar, simplemente, si un test «mide lo qué se supone que tiene que medir». Precisiones recientes realizadas por expertos en medición, además de lo que se define en *Standards for Educational and Psychological Testing* (1985; 2000), dan a entender claramente lo que es una inferencia, empleo, o consecuencia que es válida o inválida, no un test. Por ejemplo:

La validez se refiere a la idoneidad, significatividad y utilidad de las inferencias específicas realizadas a partir de las puntuaciones del test (Standards, 1985, p. 9).

La validez se refiere al grado en que una evidencia y teoría dan soporte a las interpretaciones de las puntuaciones del test que conllevan los usos específicos de los tests. (Standards, 2000, p. 9).

[Validez es] un juicio evaluativo integrado del grado en el que una evidencia empírica y las bases lógicas teóricas dan soporte a la adecuación e idoneidad de las deducciones y acciones basadas en las puntuaciones del test u otros modos de diagnóstico (Messick, 1989, p. 13).

Luego, para asegurar la validez, el investigador necesita identificar suposiciones o encontrar argumentos para justificar una deducción o emplearlos para un propósito específico, (por ejemplo, concluir que los estudiantes de un grupo poseen mayores conocimientos o tienen una autoestima más alta que los estudiantes de otro grupo) y luego, reunir pruebas para dar soporte a esas suposiciones (Shepard, 1993). Este énfasis es compatible con la idea de que la validez es un concepto singular, unitario, que requiere evidencia para el uso específico que se ha mencionado. De esto se deduce que un test puede ser no válido o inválido en sí mismo porque el mismo test puede utilizarse para objetivos diferentes. Por ejemplo, un test para el ingreso en la universidad puede llevar a deducciones válidas sobre el comportamiento futuro de un estudiante como estudiante de licenciatura, pero, no a deducciones inválidas sobre la calidad del programa del instituto.

En la investigación educativa se utilizan, esencialmente, dos tipos de inferencias. La primera está relacionada con la valoración de los logros, que depende principalmente de cómo el contenido de un test u otra evaluación representa un muy extenso campo de acción de contenidos o tareas. Para este tipo de inferencia, la evidencia basada en el contenido del asesoramiento, es necesario apoyar las inferencias que se establecen. Un segundo tipo de inferencia que es, incluso, más común en investigación educativa, se refiere a los rasgos o características que son más abstractos que el contenido claramente definido. Estos rasgos o características, a menudo, se denominan *constructos*, e incluyen, por ejemplo, inteligencia, creatividad, comprensión lectora, actitudes, razonamiento y autoestima.

Cuando las inferencias implican estos constructos, es importante tener una conceptualización teórica clara de lo que se está midiendo y la evidencia de que no existe una hipótesis contraria factible que modifique la interpretación deseada. Pueden considerarse dos tipos de hipótesis contrarias, *constructo infraestimado* y *varianza residual del constructo*. El **constructo infraestimado** acontece si el asesoramiento falla al capturar aspectos importantes del constructo. Por ejemplo, si la medida de la autoestima no incluía cuestiones sobre aspectos sociales ni académicos, mediría menos que el constructo «autoestima» propuesto.

La **varianza residual del constructo** se refiere al grado en el que una medida incluye material o factores que resultan extraños al constructo previsto. Un ejemplo de este tipo de factor sería la medición del razonamiento matemático con problemas de narraciones. Como es necesaria la comprensión lectora para entender los problemas, esta capacidad es tan importante como el razonamiento matemático para el éxito. De este modo, la medida está influenciada, en cierta medida, por factores que no forman parte del constructo.

Si la inferencia implicada en la investigación es un contenido primario o constructo, existen cinco tipos principales de evidencia que pueden usarse tanto para apoyar las interpretaciones previstas como para eliminar cualquier hipótesis contraria sobre lo que se está midiendo: la evidencia basada en el contenido, en los procesos de respuesta, en la estructura interna, en las relaciones con otras variables y en las consecuencias. Consideraremos las cuatro primeras, que poseen mayor relevancia para la investigación.

La evidencia basada en el contenido del test. En general, **la evidencia basada en el contenido del test** demuestra el grado en que la muestra de ítems o cuestiones de la prueba es representativo de algún universo apropiado o ámbito de contenidos o tareas. Este tipo de evidencia se acumula, por lo general, por expertos que examinan los contenidos de la prueba e indican la escala con la que miden objetivos o criterios predeterminados. Los expertos también juzgan la criticidad relativa o importancia de las diversas partes de la prueba. Por ejemplo, para aumentar la evidencia de un test de conocimiento sobre profesores eventuales, es necesario tener expertos que examinen los ítems y juzguen su representatividad (por ejemplo, ¿es representativa una pregunta sobre Piaget de lo que se necesita conocer sobre el desarrollo del niño?) y si el porcentaje del test dedicado a los diferentes asuntos es apropiado (por ejemplo, 20% del test se refiere a la gestión del aula, pero puede ser que debería ser un 40%). La evidencia basada en el contenido del test es esencial para tests de rendimiento. Además, el ámbito o universo que se representa debería ser apropiado al empleo que se pretende dar a los resultados.

Desafortunadamente, la validez de la evidencia basada en el contenido del test no se presenta en los artículos de investigación, porque normalmente no hay un esfuerzo para obtener tal evidencia a partir de las pruebas diseñadas de forma limitada. Cuando se recurre a pruebas estandarizadas, es importante referirse a investigaciones previas, a revisiones del procedimiento utilizado o a manuales técnicos.

La evidencia basada en el contenido es similar a la *validez aparente*, pero la validez aparente es una estimación menos sistemática entre la medida y el ámbito más extenso. La validez aparente es un juicio en el que los ítems parecen ser relevantes, mientras que la validez de la evidencia del contenido establece la relación de forma empírica.

La evidencia basada en los procesos de respuesta. **La evidencia basada en los procesos de respuesta** se centra en un análisis de las estrategias de ejecución o

de las respuestas a tareas específicas, y si estas estrategias y respuestas concuerdan con lo que se pretende medir. Por ejemplo, si los estudiantes deben conocer un razonamiento matemático, entonces sería posible preguntarles acerca de su manera de pensar en relación con la resolución de problemas para aplicar este razonamiento, y no se recurre a la aplicación maquinal de un algoritmo. De forma similar, se puede solicitar a los observadores o jueces que indiquen los criterios que han utilizado para sus juicios con el fin de estar seguros que se han aplicado los criterios apropiados.

La evidencia basada en la estructura interna. La estructura interna de un instrumento o prueba se refiere a la relación que mantienen los ítems entre sí y cómo se relacionan las diferentes partes de una prueba. La **evidencia basada en la estructura interna** se proporciona cuando las relaciones entre los ítems y entre las partes de la prueba son compatibles, experimentalmente, con la teoría o el empleo deseado de las puntuaciones. De este modo, si una medida de autoestima postula varios «tipos» de autoestima (por ejemplo, académica, social, atlética), entonces los ítems de medición del componente académico deberían estar fuertemente relacionados entre sí y no tan fuertemente relacionados con los otros componentes.

La evidencia basada en las relaciones con otras variables. La forma más frecuente de establecer la validez de las interpretaciones es mostrar cómo las puntuaciones de una medida dada se relacionan tanto con rasgos similares como, también, con rasgos diferentes. Existen varias formas para poder hacerlo. Cuando las puntuaciones de una prueba se relacionan fuertemente con las puntuaciones de otra medida del mismo rasgo, tenemos lo que se denomina *validez convergente*. La validez discriminante existe cuando las puntuaciones no se relacionan en gran medida con las puntuaciones de una prueba que mide algo diferente. De esta manera, podríamos esperar que las puntuaciones de una medida de autoestima se correspondieran totalmente con otras medidas de autoestima y mostrarán menos correlación, sin embargo, con otros rasgos diferentes como por ejemplo, ansiedad y comportamiento académico. En muchos artículos de investigación, este tipo de validez se denominará «validez de constructo».

Otra aproximación al acopio de **evidencia basada en las relaciones con otras variables** corresponde al grado en que las puntuaciones o medidas del test predicen la actuación sobre una medida de criterio (relaciones test-criterio). Se recurre a dos recomendaciones para obtener una validez test-criterio: predictiva y concurrente. Con la validez predictiva, el criterio se mide a la vez en el futuro, después de que se haya administrado la prueba. La evidencia vincula cómo la medida inicial puede predecir el criterio de conducta o su ejecución. Por ejemplo, en la recogida de evidencia sobre una nueva medida para seleccionar aspirantes para puestos de dirección, las puntuaciones de la prueba se relacionarían con el comportamiento de liderazgo futuro. Si las personas que puntuaron bajo en el test resultaron ser malos directivos y los que puntuaron alto fueron buenos directivos, podría

obtenerse validez predictiva de test-criterio. Con la validez concurrente, la prueba y el criterio se alcanzan, más o menos, al mismo tiempo. La validez referida al criterio, a menudo, se documenta en la investigación con la indicación de que una medida se relaciona con un criterio que asegura la misma cosa. Por ejemplo, en el extracto 6.1, a partir de un estudio publicado sobre el comportamiento social de un estudiante en la clase, los investigadores presentaron un acuerdo entre las percepciones de los observadores y profesores sobre comportamientos sociales.

La validez es claramente el aspecto individual más importante de una prueba y de los hallazgos que se obtienen de los datos. La calidad de la validez de los hallazgos juzgada por los usuarios varía mucho en la investigación educativa. Si se utilizan los tests estandarizados se producirá una evidencia alterada, mientras los cuestionarios desarrollados de forma limitada pueden tener una pequeña validez sistemática. En cualquier caso, los investigadores rigurosos siempre se preguntan: ¿son apropiadas las conclusiones?, ¿qué validez apoya mi conclusión? Los componentes del test de validez se resumen en la tabla 6.1.

EXTRACTO 6.1

VALIDEZ BASADA EN LAS RELACIONES CON OTRAS VARIABLES

La medida de cooperación del estudiante en el cuestionario del profesor mostró una significativa correlación positiva ... con la medida de observación del comportamiento colaborador y del comportamiento amistoso ... El test sobre desorganización del estudiante puntuó de forma negativa cuando se puntuó el comportamiento deportivo y amistoso ... la colaboración entre estudiantes y ... la sociabilidad espontánea.

Fuente: Slavin, R. E. (1979), «Effects of biracial learnings teamson cross-racial friendships», *Journal of Educational Psychology*, 71. Reimpreso con permiso del autor.

Efecto de la validez sobre la investigación. Debido a que la validez implica una interpretación apropiada y la utilización de la información recogida a través de la medición, es necesario tanto para usuarios como para investigadores de la investigación juzgar el grado de validez que se presenta, basado en la evidencia disponible. En este sentido, la validez es una cuestión de grado y no un planteamiento de todo o nada. Los investigadores deberían mostrar que las deducciones y conclusiones específicas de sus estudios poseen una evidencia para la que existe validez. Los usuarios necesitan tomar la misma decisión basada en su empleo de los resultados. ¿Se sugiere que la validez debe establecerse para cada situación de investigación y cada empleo posible? Tal requerimiento añadiría una considerable cantidad de recogida de datos y de análisis a cada estudio y, por lo tanto, lo haría irrealizable. En la práctica, es necesario generalizar a partir de otros estudios y de otras investigaciones cuya

TABLA 6.1: COMPONENTES DE LA VALIDEZ DEL TEST

Componente	Descripción	Prueba
Evidencia basada en el contenido.	Grado en el que los ítems o cuestiones representan un amplio dominio.	Examinar la relación entre el contenido de los ítems y el contenido del dominio.
Evidencia basada en los procesos de respuesta.	Si los pensamientos y los procesos de respuesta son acordes con la interpretación deseada.	Examinar las explicaciones del encuestado y los patrones de respuesta.
Evidencia basada en la estructura interna.	Grado en el que los ítems miden el mismo rasgo relacionado.	Correlacionar ítems de medida del mismo rasgo.
Evidencia basada en las relaciones con otras variables.	Si la medida se relaciona con variables similares o predichas y no se relaciona con variables diferentes.	Correlacionar la medida con otras medidas del mismo rasgo y con medidas de rasgos diferentes.

interpretación y empleo resulten válidos. Esta es una razón por la que las pruebas establecidas, para las cuales cualquier evidencia de validez debe ser probablemente acumulada, por lo general, proporcionan medidas más fiables. De otra manera, sería un error asumir que sólo porque una prueba está establecida, sus resultados son válidos.

Las pruebas diseñadas de forma limitada, sin historia de uso o revisiones efectuadas por otros, necesitan evaluarse con más cuidado. Cuando los investigadores desarrollan un nuevo procedimiento, es muy importante reunir una evidencia apropiada para la validez y, luego, presentar esta evidencia en el estudio.

Si se utiliza una prueba preparada o establecida de forma limitada, lo mejor es reunir una evidencia para la validez antes de que se recojan los datos para el estudio. Esta es la principal razón para un test piloto de cualquier prueba y de las técnicas para administrarla. El investigador debería ser consecuente con el empleo de los resultados. Por ejemplo, si empleará los resultados para determinar qué estudiantes poseen los conocimientos básicos, es necesaria la validez del contenido. Si está examinando una teoría relacionada con el desarrollo de un estudio cognitivo, se necesita la validez del constructo.

Fiabilidad

La **fiabilidad** se refiere a la coherencia de la medición, el grado en el que los resultados son similares sobre formularios diferentes de la misma prueba o de las circunstancias de la recogida de datos. Otra forma de conceptualizar la fiabilidad

es determinar el grado en que las medidas están libres de error. Si una prueba tiene un pequeño error, resulta fiable, y si tiene una gran cantidad de errores, es de poca confianza. Podemos medir el error mediante la estimación de la coherencia de la valoración de un rasgo.

Piense por un minuto en los tests que ha realizado. ¿Las puntuaciones que recibió eran exactas o encerraban algún grado de «error» en los resultados? ¿Eran algunos resultados más exactos que otros? En la medición de rasgos humanos, como rendimiento, actitud, personalidad, estado físico o cualquier otro rasgo, casi nunca obtendrá unos resultados que no posean algún grado de error. Muchos factores contribuyen a desnaturalizar nuestras medidas. Puede haber preguntas ambiguas, la iluminación puede ser escasa, algunos sujetos pueden estar enfermos, la estimación de un test de rendimiento puede ser afortunada o desafortunada, los observadores pueden estar cansados y así sucesivamente. Lo que esto significa es que, incluso, cuando un rasgo permanecía constante al realizar dos tests con una semana de diferencia, las puntuaciones podrían no ser exactamente las mismas debido a un error inevitable.

La puntuación obtenida puede considerarse como si tuviéramos dos componentes, *uno verdadero* o *universo de puntuación*, que representa el conocimiento actual o el nivel de capacidad del individuo, y *otro erróneo*, fuentes de variabilidad no relacionadas con el propósito de la prueba:

$$\text{puntuación obtenida} = \text{universo de puntuación o verdadero} + \text{error}$$

En la tabla 6.2 se enumeran las fuentes de error más comunes. Por tanto, el objetivo de la selección o evaluación de la prueba, es buscar la evidencia de que el error se ha controlado lo más posible.

TABLA 6.2: FUENTES DE ERROR DE MEDICIÓN

Condiciones de la administración y elaboración del test	Circunstancias asociadas a la persona a la que se aplica el test
Cambios en los límites de tiempo Cambios en las instrucciones Pruebas de puntuación diferentes Sesión de comprobación interrumpida Raza del examinador Tiempo disponible para el test Ejemplos de ítems Ambigüedad en la redacción Incorrecta interpretación de las instrucciones Efectos del calor, luz, ventilación en la sala del test Diferencias entre los observadores	Reacciones a ítems específicos Salud Motivación Humor Fatiga Suerte Fluctuación en la memoria o atención Actitudes Niveles del test Capacidad para la comprensión de instrucciones Ansiedad

La cantidad actual de varianza de error en las puntuaciones de los tests, o la fiabilidad, se determina empíricamente a través de varios tipos de procesos¹. Cada tipo de fiabilidad se relaciona con el control de un tipo determinado de error y, normalmente, se presenta en forma de coeficiente de fiabilidad. El coeficiente de fiabilidad es una estadística de correlación que compara dos conjuntos de puntuaciones a partir de los mismos individuos. La escala del coeficiente de fiabilidad abarca de .00 a .99. Si el coeficiente es alto, por ejemplo 0.90, las puntuaciones tienen un error pequeño y son altamente fiables. Lo contrario también es cierto para la correlación cercana a .20 o .35. Un rango aceptable de fiabilidad para los coeficientes de la mayoría de las pruebas va desde .70 hasta .90.

Los cinco tipos generales de estimación de la fiabilidad son la estabilidad, la equivalencia, la estabilidad y la equivalencia, la consistencia interna y el acuerdo (tabla 6.3)².

La estabilidad. Un coeficiente de **estabilidad** se obtiene por la correlación de puntuaciones a partir del mismo test realizado a un grupo de individuos en dos ocasiones diferentes. Si las respuestas de los individuos son coherentes (esto es, si aquello que puntuó alto la primera vez, sigue puntuando alto la segunda vez, y así sucesivamente), entonces, el coeficiente de correlación y la fiabilidad son altos. Este procedimiento test-retest asegura que las características que se han medido permanecen constantes. Los rasgos inestables, como por ejemplo, el humor, no se espera que tengan coeficientes de estabilidad altos. Además, la estabilidad, normalmente significa que hay tiempo suficiente entre las medidas (a menudo, varios meses) para que la coherencia en las puntuaciones no se vea influida por una memoria o efecto práctico. En general, cuando el espacio de tiempo entre las mediciones se incrementa, la correlación entre las puntuaciones disminuye.

La equivalencia. Cuando dos formas equivalentes o paralelas de la misma prueba se administran a un grupo en el mismo momento, y las puntuaciones están rela-

¹ La mayoría de procedimientos se basan en la suposición de que habrá suficiente dispersión o amplitud en las puntuaciones para calcular los coeficientes de correlación. Algunos tipos de test (como por ejemplo, los de referencia criterial) no proporcionan mucha variabilidad de puntuación y los indicadores correlacionales tradicionales de fiabilidad pueden resultar inapropiados. Para estos test, los investigadores examinan los porcentajes de los examinados que se clasifican de la misma forma después de realizar el test por segunda vez o después de aplicar formatos diferentes del mismo test; o el porcentaje de respuestas que son iguales en momentos diferentes, además del coeficiente de correlación. La presentación de la fiabilidad en este capítulo se enfocará desde los procedimientos correlacionales tradicionales, debido a que son los únicos que encontrará de forma más frecuente en la bibliografía.

² De acuerdo con los nuevos *Standards for Educational and Psychological Testing* estos índices de fiabilidad tradicionales son casos especiales de una clasificación más general denominada *teoría de la generabilidad*. La teoría de la generabilidad tiene la habilidad de combinar varias fuentes de error en una medida de variabilidad única. Y, aunque, normalmente no aparece en la bibliografía, proporciona una indicación más exacta del grado de error.

TABLA 6.3: TIPOS DE FIABILIDAD

Tipo	Descripción	Procedimiento	Ejemplos habituales ^a
Estabilidad	Coherencia de las características estables a través del tiempo.	Administrar el mismo test a los mismos individuos a través del tiempo.	Test de aptitud. Test IQ.
Equivalencia	Equiparabilidad de dos medidas del mismo rango en el mismo momento.	Administrar formularios diferentes a los mismos individuos al mismo tiempo.	Test de rendimiento.
Equivalencia y estabilidad	Equiparabilidad de dos medidas del mismo rango a través del tiempo.	Administrar formularios diferentes a los mismos individuos a través del tiempo.	Valoración de los cambios a través del tiempo. Valoración de la personalidad.
Consistencia interna (por dos mitades, K-R, alfa Cronbach)	Equiparabilidad por dos mitades de una medida para asegurar un rasgo concreto o dimensión.	Administrar un test y correlacionar los ítems unos con otros.	La mayoría de las medidas, excepto para test de velocidad. Cuestionarios de actitud.
Acuerdo	Coherencia de valoraciones u observaciones.	Dos o más personas valoran u observan.	Observaciones y entrevistas.

^a Estos ejemplos no sugieren que otras formas de fiabilidad, además de las indicadas, sean inapropiadas (por ejemplo, los test de rendimiento también emplean test-retest de fiabilidad).

cionadas, la fiabilidad resultante es un coeficiente de **equivalencia**. Aunque cada forma se hace para cuestiones diferentes, la puntuación de cada individuo sería la misma en ambas formas. La equivalencia es un tipo de fiabilidad que puede establecerse cuando el investigador tiene un número relativamente grande de ítems a partir de los cuales construye formas equivalentes. Las formas alternativas de un test son necesarias para comprobar, inicialmente, los sujetos ausentes que pueden aprender sobre los ítems específicos de la primera forma o cuando un profesor tiene dos o más secciones de la misma clase en momentos diferentes.

La equivalencia y la estabilidad. Cuando un investigador necesita hacer un examen preliminar y otro posterior para asegurar un cambio en el comportamiento,

debe establecerse un coeficiente de equivalencia y estabilidad. En este proceso, los datos de fiabilidad se obtienen mediante la administración, al mismo grupo de individuos, de una forma de una prueba en un momento dado, y una segunda forma un tiempo después. Si una prueba tiene este tipo de fiabilidad, el investigador puede estar seguro de que un cambio en las puntuaciones a través del tiempo, refleja una diferencia vigente en el rasgo que se mide. Este es el tipo de fiabilidad más rígido y resulta especialmente útil para estudios que incluyen puntuaciones de ganancia o mejora.

La consistencia interna. La **consistencia interna** es el tipo de fiabilidad más frecuente desde que puede estimarse a partir de la entrega, única, de una forma de un test. Existen tres tipos comunes de consistencia interna: por dos mitades, *Kuder-Richardson*, y el método alfa de *Cronbach*. En la fiabilidad **por dos mitades**, los ítems de un test que se han administrado a un grupo se dividen en dos mitades comparables, y el coeficiente de correlación se calcula entre las mitades. Si cada estudiante tiene aproximadamente la misma posición en su relación con el grupo en cada mitad, entonces, la correlación es alta y el instrumento es altamente fiable. Cada mitad del test debería tener una dificultad similar. Este método proporciona una fiabilidad más baja que otros métodos, ya que el número total en la ecuación de correlación contiene sólo la mitad de los ítems (y sabemos que todo lo demás es igual, los test más largos son más fiables que los cortos). (La fórmula *Spearman-Brown* se utiliza para incrementar las fiabilidades entre dos mitades para estimar cuál sería la correlación para un test completo). Las técnicas de consistencia interna no deberían utilizarse con tests rápidos. Esto se debe a que los estudiantes no contestan todos los ítems, un factor que tiende a incrementar falsamente la correlación de los ítems.

Un segundo método para la investigación del grado de consistencia interna es emplear la fórmula **Kuder-Richardson** (K-R) para correlacionar todos los ítems de un test único, unos con otros, cuando cada ítem puntúa correcta o incorrectamente. La fiabilidad K-R de hecho, se determina a partir de la única administración de una prueba, pero sin la necesidad de dividirla en dos mitades equivalentes. Este procedimiento asegura que todos los ítems de una prueba son equivalentes unos con otros, y que es apropiado cuando el propósito del test es medir un único rasgo. Si un test posee ítems de varias dificultades o si mide más de un rasgo, la estimación K-R normalmente será menor que la fiabilidad por dos mitades.

El **alfa Cronbach** también supone la equivalencia de todos los ítems. Es una forma de consistencia interna mucho más general que la K-R, y se utiliza para ítems que no han puntuado correcta o incorrectamente. El alfa Cronbach normalmente es el tipo de fiabilidad más apropiado para investigación de estudio y otros cuestionarios en los que hay un rango de posibles respuestas para cada ítem.

El acuerdo. El quinto tipo de fiabilidad se expresa como un coeficiente de **acuerdo**. Se establece determinando el grado en el que dos o más personas están

de acuerdo sobre lo que han visto, escuchado o evaluado. Esto es, cuando dos o más observadores o evaluadores observan o evalúan independientemente algún aspecto concreto, ¿estarán de acuerdo entre sí sobre lo que han observado o evaluado? Si es así, entonces, existe algo de consistencia en la medida. Este tipo de fiabilidad se emplea, normalmente, para investigación observacional y en estudios basados en comportamientos en los que los juicios profesionales se hacen sobre el comportamiento del estudiante. Se presentará como fiabilidad inter-rater o *acuerdo tanteador* y se expresará como un coeficiente de correlación o como un porcentaje de acuerdo. Sin embargo, este tipo de análisis no indica nada sobre la *consistencia* de comportamiento o conducta en momentos diferentes (irónicamente, las estimaciones de consistencia interna no las hace cualquiera). Esto significa que es un medio para obtener un acuerdo inter-rater alto, que es relativamente sencillo de hacer, y además para obtener datos que muestren que el rasgo o comportamiento es consistente a través del tiempo.

Hemos resumido estos cinco tipos de fiabilidad mostrados en la tabla 6.4 de acuerdo a las diferentes formas en las que se suministra un instrumento.

Interpretación de coeficientes de fiabilidad. Deben considerarse varios factores en la interpretación de los coeficientes de fiabilidad:

1. Cuanto más heterogéneo sea un grupo en lo referente al rasgo que se mide, más aumentará la fiabilidad.
2. Cuantos más ítems posea una prueba, más aumentará la fiabilidad.
3. Cuanto mayor sea el rango de puntuaciones, más aumentará la fiabilidad.

TABLA 6.4: PROCEDIMIENTOS PARA LA ESTIMACIÓN DE LA FIABILIDAD ^a		
	Tiempo 1	Tiempo 2
Estabilidad	A	A
Equivalencia	A B	
Estabilidad y equivalencia	A	B
Consistencia interna	A	
Acuerdo	R1 R2	

^a A y B se refieren a diferentes formas del mismo test; R1 y R2 se refieren al evaluador 1 y evaluador 2, sin embargo, pueden usarse más de dos evaluadores u observadores con el acuerdo.

Fuente: Adaptación de McMillan, J. H. (2000), *Educational Research: Fundamentals for the Consumer* (3.ª ed.), New York: Longman.

4. Los test de rendimiento con un nivel medio de dificultad tendrán una fiabilidad más alta que cualquier otro test más fácil o más difícil.

5. La fiabilidad, al igual que la validez, cuando se basan en un grupo matinal, se demuestra sólo para sujetos con características similares a los del grupo matinal.
6. Mientras más discriminen los ítems entre las personas con mucho éxito, mayor será la fiabilidad.

Los investigadores, a menudo, preguntan cómo una correlación alta podría indicar una fiabilidad satisfactoria. Esta pregunta no tiene una respuesta sencilla. Depende del tipo de prueba (los cuestionarios de personalidad, generalmente, tienen una fiabilidad menor que los tests de rendimiento), del propósito del estudio (si es investigación exploratoria o investigación que conduce a decisiones importantes), y si los resultados afectan a los individuos o grupos (ya que la acción que afecta a los individuos requiere más correlación que la acción que afecta a los grupos). Sin embargo, una acertada regla general es tener cuidado con fiabilidades por debajo de .70.

Efecto de la fiabilidad en la investigación. Al igual que sucede con la validez, la fiabilidad de las puntuaciones debería establecerse antes de que el investigador aborde la investigación, y el tipo de fiabilidad debería ser consistente con el uso de los resultados. Si va a utilizar los resultados para selección o predicción en programas especiales, se necesitan estimaciones estables de fiabilidad. Si está interesado en programas para modificar actitudes o valores, se necesitan estimaciones de equivalencia. La fiabilidad también debe establecerse con individuos que sean parecidos a los sujetos de la investigación. Si los estudios previos presentan una elevada fiabilidad con alumnos de secundaria y se pretende emplear los resultados con alumnos de primaria, la fiabilidad puede no ser adecuada. Más comúnmente, la fiabilidad se presenta con los alumnos utilizados en el estudio. Un fallo en la presentación de la fiabilidad sería causa de interpretar los resultados con precaución; sin embargo, hay algunas medidas sencillas para las que los coeficientes de fiabilidad no son necesarios (ver a continuación).

Leerá alguna investigación en la que la fiabilidad no esté consignada, pero los resultados de la investigación muestran lo que se llama «diferencias significativas». Esta es una situación interesante en investigación, debido a que es más difícil encontrar diferencias entre grupos mediante las pruebas de las que se obtienen puntuaciones que tienen una fiabilidad baja. Es como si las diferencias se observaran a pesar de que tienen una fiabilidad baja. Por supuesto, es posible que la medición fuese fiable, aunque no se presentaran estimaciones fiables. Esta situación es más probable que ocurra en investigaciones en la que los sujetos responden a cuestiones tan sinceras y simples que la fiabilidad se «da por supuesta». Por ejemplo, en los estudios de las percepciones sobre éxito o fracaso de los estudiantes, a los sujetos se les pidió que indicaran sobre una escala de 1 a 10 (1 grado más alto de fracaso y 10 grados más altos de éxito) sus sensaciones de éxito o fracaso. En muchas investigaciones, los sujetos presentan información como

por ejemplo, edad, sexo, ingresos, tiempo empleado en estudiar, profesión, y otras cuestiones relativamente sencillas. Para estos tipos de datos, normalmente, no se necesitan estimaciones estadísticas de la fiabilidad.

La fiabilidad es una función de la naturaleza del rasgo que se mide. Algunas variables, como por ejemplo, la mayoría de las medidas de rendimiento, proporcionan puntuaciones altamente fiables, mientras que las puntuaciones de medidas de personalidad tienen fiabilidades menores. Consecuentemente, para la medición de rasgos de personalidad una fiabilidad de .70 puede ser aceptable. Entonces, en comparación, una prueba de personalidad que presenta un coeficiente de personalidad de .90 sería considerada con una fiabilidad excelente, y un test de rendimiento con una fiabilidad de .70 se vería como poco fiable. Necesitamos una fiabilidad mucho más fuerte si los resultados van a usarse para tomar decisiones sobre los individuos. Los estudios de grupos pueden tolerar una fiabilidad menor, algunas veces tan bajas como .50 en investigación exploratoria. Las medidas sobre niños pequeños, normalmente, son menos fiables que aquéllas sobre sujetos con más edad.

Para incrementar la fiabilidad, lo mejor es establecer condiciones estándar de recogida de datos. Se deberían dar las mismas instrucciones a todos los sujetos, deberían tener el mismo tiempo para contestar las preguntas y a la misma hora del día, y así sucesivamente. Normalmente se incrementa el error si la prueba la administran diferentes personas. Es importante conocer si tiene lugar alguna circunstancia extraordinaria durante la recogida de datos, ya que puede afectar a la fiabilidad. Para ser fiables, las pruebas tienen que ser apropiadas al nivel de lenguaje y lectura, y debe motivarse correctamente a los sujetos para que contesten a las preguntas. En alguna investigación, es difícil encontrar sujetos que sean sinceros, por ejemplo, cuando se les pide a los estudiantes que hagan test de rendimiento en los que no se sienten implicados. La fiabilidad también puede sufrir cuando se solicita a los sujetos que completen varias pruebas muy largas. Una hora es más o menos lo que cualquiera de nosotros puede soportar, y para los niños más pequeños el máximo es menos de media hora. Si se dan varias pruebas al mismo tiempo, el orden en que se administran no debería ser el mismo para todos los sujetos. Algunos sujetos contestarían una prueba primero y otros sujetos contestarían la misma prueba al final. Esto se denomina *contrabalancear* las pruebas. Si se dan varias pruebas y no hay contrabalanceo, los resultados, especialmente de las pruebas que se realizan al final, deberían observarse con más precaución.

Finalmente, la fiabilidad es una condición necesaria para la validez. Esto es, las puntuaciones no son válidas a no ser que sean fiables. Sin embargo, una medida fiable no es necesariamente válida. Por ejemplo, podemos obtener una medida muy fiable de la longitud de su dedo gordo del pie, ¡pero no será válida como una estimación de su inteligencia!

En el resto de este capítulo consideraremos los métodos de recogida de datos que se usan más comúnmente en investigación cuantitativa. Incluso, cuando los principios básicos de la validez y la fiabilidad aplicados a los cinco tipos de recogida de datos, hay que darse cuenta de que cada técnica de recogida de datos posee características únicas que afectan a la forma en que se establece la validez y la fiabilidad.

PRUEBAS DE LÁPIZ Y PAPEL

El término «prueba de lápiz y papel» significa que se presenta una batería estándar de preguntas por escrito a cada sujeto (en papel u ordenador) que requiere la realización de tareas cognitivas. Las respuestas o contestaciones se resumen para obtener un valor numérico que representa una característica del sujeto. La tarea cognitiva puede enfocarse sobre lo que la persona conoce (logro), es capaz de aprender (capacidad o aptitud), elecciones o selecciones (intereses, actitudes o valores), o es capaz de hacer (habilidades). En este capítulo, se resumen brevemente los diferentes tipos de tests y sus usos en la investigación, pero es preciso destacar que todos los tests miden comportamientos actuales. Los tests difieren más en su empleo que en su desarrollo o en sus ítems, particularmente, cuando comparamos tests de aptitud y rendimiento. De hecho, sería más exacto decir que existen diferentes tipos de deducciones y usos, es lo que usted hace con los resultados experimentales que crean distinciones como el logro y la aptitud.

Tests estandarizados

Los **tests estandarizados** proporcionan procedimientos uniformes para la administración y puntuación de la prueba. Cada vez que se aplica el test, se pide que se realicen las mismas cuestiones, con un conjunto de instrucciones que especifican como debe administrarse el test. Esto incluiría información sobre las aptitudes de la persona que administra el test y las condiciones en las que debe administrarse, como por ejemplo: el tiempo permitido, materiales que pueden usarse por los sujetos, y si pueden aclararse cuestiones sobre el test durante su realización. Normalmente, la puntuación de las respuestas es objetiva, y la mayoría, pero no todos los tests estandarizados se realizan en grupos matinales. El grupo tipo, como se denomina, permite la comparación de una puntuación con la realización de un grupo definido de individuos. Esto proporciona información importante y valiosa, pero el investigador debe tener cuidado al interpretar las puntuaciones referidas a normas (ver sección a continuación: pruebas con referencia a criterio y a normas).

La mayoría de los tests estandarizados se preparan comercialmente por expertos en medición. Esto significa que se prestará una atención especial a la naturaleza de

las normas, la fiabilidad y la validez. Los resultados de la prueba son objetivos y están relativamente poco influidos por la persona que administra la prueba. Debido a que la mayoría de los tests estandarizados se preparan comercialmente, se pretende que puedan usarse en una amplia variedad de ambientes. Consecuentemente, cualquier suceso que se mida, está típicamente definido en términos generales. Esto puede significar que para algunos propósitos de investigación un test estándar no sea lo suficientemente específico como para proporcionar una medida sensible de la variable. Por ejemplo, si está dirigiendo un estudio para investigar el efecto de la educación general en la universidad sobre el conocimiento en sociología o humanidades de los estudiantes, un test estandarizado que usted pudiera utilizar se podría pretender que fuera una medida de los resultados de sociología y de humanidades en casi todas las universidades. Esto significa que lo que se enseña en una universidad determinada no estaría bien representado en el test. Esto muestra un intercambio en el uso de tests estandarizados en la investigación. Por un lado, se tiene una prueba elaborada con mucho cuidado, con fiabilidad establecida, instrucciones y procedimientos de puntuación. Por otro lado, el test no puede enfocarse directamente sobre la variable de interés en el estudio, puede tener normas inapropiadas o resultar muy costoso. La alternativa es desarrollar su propia prueba que medirá la variable con más precisión, aunque puede tener dudosas cualidades técnicas.

Interpretación con referencia a criterios y a normas

Una diferencia principal entre los tests es si poseen referencia por criterio o referencia a normas. El propósito de una **interpretación referida a normas** es mostrar como las puntuaciones individuales se comparan con las puntuaciones de un grupo de referencia bien definido o grupo tipo de individuos. Entonces, la interpretación de los resultados depende por completo de cómo unos sujetos se comparan con otros, con menor énfasis sobre la cantidad absoluta de conocimiento o habilidad. Esto es, lo que tiene más importancia es el grupo de comparación y la habilidad del test o de la prueba para distinguir entre los individuos. La meta es saber si, por ejemplo, los sujetos saben más o menos que el grupo tipo y, a menudo, se presenta la puntuación para indicar específicamente donde «está» el sujeto en relación con los demás (como por ejemplo el percentil 67°, o el cuartil superior).

Los investigadores necesitan tener en cuenta dos características de las interpretaciones referidas a normas. Primero, debido a que el propósito del test es diferenciar entre los individuos, la mejor distribución de puntuaciones es la que muestre una varianza alta. Para lograr una variabilidad de puntuaciones alta, deben discriminarse los ítems entre los individuos. Para lograrlo, los ítems del test, particularmente en tests estándar referidos a normas, son bastante difíciles. No es raro que estudiantes en el percentil 50° o 60° respondan poco más de la mitad de la

preguntas correctamente. Los ítems fáciles, los que casi todo el mundo contesta correctamente, se utilizan con moderación (obviamente, si todos los ítems son fáciles, todo el mundo obtendría la máxima puntuación, y no habría diferencia entre los individuos). De esta manera, el contenido importante o las habilidades no pueden medirse, lo que afectará al significado que le demos a los resultados. Como aspecto positivo, la variabilidad amplia ayuda a establecer relación. Las mayores correlaciones, a menudo, se encuentran con dos variables que tienen una gran variabilidad.

Segundo, los investigadores deben centrar toda su atención en las características de grupo tipo o de referencia. Tal vez, usted ha tenido la misma experiencia que nosotros, matriculado en una clase de estudiantes listos, algunos estudiantes muy trabajadores con un instructor que les calificó por la curva. Usted podría aprender mucho pero seguía obteniendo notas bajas. La interpretación de las puntuaciones referidas a normas sólo tiene sentido cuando entendemos qué es lo que estamos comparando. Muchos test referidos a normas indican que usan normas de carácter «nacional». A pesar del hecho de que el término «nacional» está sujeto a diferentes interpretaciones, si está observando a estudiantes superdotados y compara sus puntuaciones con la norma nacional, las oportunidades son buenas, todos sus estudiantes puntuarán muy alto y mostrarán una variabilidad baja. Esto le da lo que se llama *efecto techo* y un rango restringido, que a su vez le llevará a resultados no significativos.

En una interpretación con **referencia a criterios** o **basada en patrones**, la puntuación de un individuo se interpreta por la comparación de la puntuación con patrones de conducta valorados profesionalmente. La comparación es entre la puntuación y un criterio o patrón, más que con las puntuaciones de los demás. El resultado, normalmente, se expresa como el porcentaje de ítems contestados correctamente, o como acierto-fallo en el caso de comprobación de competencia mínima. Hay un enfoque en «qué» los sujetos pueden hacer, con una comparación de ello con normas de destreza. La mayoría de los tests con referencia a criterio tienen como resultado una distribución altamente sesgada, que disminuye la variabilidad. A pesar de esta limitación, los tests con referencia a criterio son adecuados para su utilización en diagnósticos y para clasificar sujetos en grupos de acertar-fallar. Un tipo de test relacionado, con referencia a dominio, se emplea para mostrar cuánto del gran «dominio» o conocimiento específicamente definido se demuestra por aquellos que lo comprueban. Por ejemplo, si el dominio es el conocimiento o la suma con números de tres cifras, el test comprobará este dominio, y el investigador realizará un juicio profesional utilizando el porcentaje de ítems contestados correctamente para juzgar la «maestría» del dominio.

Tests de aptitud

El propósito de un **test de aptitud** es predecir un comportamiento futuro. Los resultados se utilizan para realizar una predicción acerca del comportamiento en

relación con varios criterios (calificaciones, efectividad de la enseñanza, certificación o puntuaciones de test) antes de la instrucción, colocación o entrenamiento. El término *aptitud* se refiere al uso previsible de las puntuaciones de un test, más que a la naturaleza de los ítems del test. Algunos términos, como por ejemplo *inteligencia* o *capacidad*, se emplean de forma indistinta con aptitud. Los tests de inteligencia se utilizan para proporcionar una medida muy general, referida normalmente a una puntuación global del test. Debido a que son muy generales, los test de inteligencia son útiles en el vaticinio de una amplia variedad de tareas. La inteligencia se mide mediante un test individual o de grupo. Para muchos investigadores, los tests de inteligencia de grupo son más adecuados y cuestan mucho menos que los tests individuales. La mayoría de los tests de grupo se diseñan para que los investigadores necesiten preparación para administrarlos y puntuarlos. Normalmente, estos tests obtienen tres puntuaciones: una de lenguaje verbal, una no verbal o puntuación de comportamiento, y una puntuación combinada. Debido a que, virtualmente, todos los niños han realizado un test de inteligencia en grupo, las puntuaciones están disponibles para su utilización en investigación con el permiso paterno o del colegio. A menudo las puntuaciones se emplean para corregir las diferencias de capacidad de grupos completos de sujetos. Algunos de los tests de inteligencia de grupo e individuales se enumeran en la tabla 6.5.

Hay muchas medidas que estiman múltiples aptitudes o tipos específicos de aptitudes. Los tests multifactoriales de aptitud se usan para proporcionar puntuaciones separadas para cada destreza o área deseada. Alguien podría argumentar que esto tiene más sentido que una puntuación única porque pueden identificarse fortalezas o debilidades relativas. Los tests multifactoriales de aptitud se han generalizado, de forma creciente, como instrumentos de asesoramiento educativo y vocacional. No obstante, la utilidad de las puntuaciones de un factor en la investigación puede resultar insegura. Debido a que puede haber unos pocos ítems que midan un factor, la fiabilidad de las puntuaciones puede ser cuestionable. El total de las puntuaciones individuales, en tanto que más generales, resulta más estable y fiable. Los tests especiales de aptitud son adecuados para la investigación, ya que se enfocan a partir de un indicio exacto de capacidad en un área. La tabla 6.5 contiene ejemplos de tests tanto de aptitud múltiple como de aptitud especial.

Pruebas de rendimiento

No siempre resultan evidentes las diferencias entre los tests de rendimiento y los tests de aptitud. A menudo, se utilizan ítems muy similares en ambos tipos de test. Sin embargo, en general, los **tests de rendimiento** tienen un alcance más limitado, están más estrechamente asociados a los sujetos escolares y miden

TABLA 6.5: EJEMPLOS DE TESTS ESTANDARIZADOS

Aptitud	Rendimiento
Inteligencia o capacidad de grupo Test de capacidad cognitiva Test de capacidad de la Escuela <i>Otis-Lenon</i> Test de evaluación escolástica	Diagnóstico Test <i>Standford</i> de comprensión matemática Test <i>Woodcock</i> de destreza lectora Test <i>Key</i> matemático de comprensión aritmética Test California de comprensión lectora
Inteligencia individual Escala de inteligencia <i>Standford-Binet</i> Escalas <i>Wechsler</i> Batería de evaluaciones para niños <i>Kaufman</i> Escalas <i>McCarthy</i> de capacidad infantil	Referidos a criterios Test y banco de ítems referidos a objetivos Sistema de monitorización de habilidades Test de destreza escritora
Multifactorial Test de aptitud diferencial Batería de tests de aptitudes generales Batería de aptitudes vocacionales de los Servicios Armados	Sujetos específicos Tests Metropolitano de destrezas Tests de lectura de <i>Gates-MacGinitie</i> Tests de comprensión de las matemáticas modernas
Especial Test eclesiástico de Minesota Test de ingreso en la Facultad de Leyes Test de ingreso en la Facultad de Medicina Test <i>Benet</i> de comprensión mecánica Test <i>Torrance</i> de pensamiento creativo Estimación <i>Watson-Glaser</i> de pensamiento crítico	Baterías Tests Terra Nova de habilidades básicas de comprensión Tests prácticos de California Tests Metropolitano de ejecución Test de habilidades básicas de Iowa Test prácticos de <i>Standford</i>

aprendizajes más actuales que los tests de aptitud. Además, por supuesto, el propósito de los tests de rendimiento es medir lo que se ha aprendido más que predecir una conducta futura.

Hay muchos tests de rendimiento estandarizados. Algunos son de diagnóstico, es decir, aíslan áreas específicas de fortalezas o debilidades; algunos tienen que ver con la medición del rendimiento en un área de contenido concreto, mientras otros (baterías de estudio) comprueban áreas de contenidos diferentes; algunos se refieren a normas y otros se refieren a criterios; algunos resaltan principios y destrezas en lugar de conocimiento de hechos específicos. La elección del test de rendimiento depende del propósito de la investigación. Si la investigación está relacionada con el rendimiento de un sujeto escolar específico, entonces, lo mejor sería recurrir a un test que midiera sólo a este sujeto, en lugar

de emplea una batería de estudio. Si se pretende hacer comparaciones entre varios colegios, lo mejor es emplear un test con referencia a normas.

Es muy importante asegurar la validez del contenido con un test de rendimiento estándar antes de usarlo en la investigación. Esto se debe a que el currículo en algunos colegios puede ser diferente del contenido de los tests estandarizados diseñados para su empleo en la mayoría de los colegios. La mejor forma de asegurar una evidencia para la validez del contenido es examinar los ítems del test y hacer juicios profesionales de la coincidencia entre lo que el ítem comprueba y el currículo. Finalmente, los que eligen los test, deberían considerar el nivel de dificultad del test y las capacidades de los alumnos. La meta deseada es poseer una evidente distribución normal de las puntuaciones del test. Si los resultados resultan sesgados por un test que es demasiado fácil o demasiado difícil para los alumnos, será difícil relacionar las puntuaciones con otras variables (como por ejemplo la medición del aumento del rendimiento a lo largo de un año). La tabla 6.5 contiene ejemplos de algunos tests de rendimiento estándar.

Pruebas alternativas

En contraste con los formatos tradicionales de comprobación que confían en ítems objetivos escritos, las **pruebas alternativas** están diseñadas para proporcionar diferentes maneras de demostrar la conducta y el rendimiento de los alumnos, a menudo, en contextos «auténticos». Confiando, fundamentalmente, en tener las respuestas de interpretación elaboradas por los estudiantes, se están empleando con creciente frecuencia pruebas alternativas, incluso, para programas de comprobación a nivel estatal. Mientras haya muchos tipos de pruebas alternativas, incluidas las demostraciones y exhibiciones, consideraremos las dos que encontraremos con más probabilidad en investigación: la basada en comportamientos y el portfolio.

Diagnósticos basados en la ejecución. Mediante el **diagnóstico basado en la ejecución**, el énfasis se sitúa en la medición de la destreza de las habilidades cognitivas del estudiante a través de la observación directa sobre cómo el estudiante ejecuta la habilidad, a menudo, en un contexto espontáneo. Los contextos son «auténticos» según el grado en que reflejan lo que harán exactamente los alumnos con lo que están aprendiendo. Por ejemplo, pueden considerarse diagnósticos basados en la ejecución de una tarea: solicitar a los alumnos que completen un proyecto de historia oral basado en un resumen de entrevistas y de fuentes escritas; escribir cartas de demandas de empleo, completar un recital de música o preparar una carpeta de ilustraciones.

Los diagnósticos basados en la ejecución de un trabajo tienen la ventaja de proporcionar una medida holística directa de las habilidades mentales que se aseguran indirectamente en los tests escritos. Estos diagnósticos, además, proporcionan

una mejor medida de la ejecución de la habilidad en los contextos más parecidos a los que los alumnos encontrarán fuera de la escuela. Los diagnósticos basados en la ejecución tienen normalmente referencia criterial, sin las restricciones, algunas veces poco realistas, arbitrarias y de tiempo de un test por escrito. Además, los diagnósticos basados en la ejecución están muy relacionados con la enseñanza. Esto significa que la investigación sobre las relaciones de las prácticas educativas con las variables dependientes que se miden por estos diagnósticos será muy susceptible, a diferencia de lo que ocurre en los tests estandarizados más generales. Sin embargo, el mayor inconveniente de los diagnósticos basados en la ejecución es la dependencia sobre evaluaciones subjetivas u observaciones de los profesores que, a menudo, ofrecen una fiabilidad baja. La mayoría de las veces, esto se convierte en un problema si se confía en una evaluación u observación concreta. Además, es necesario tener en cuenta el tiempo que supone desarrollar, aplicar a los alumnos y puntuar estos diagnósticos. Normalmente, los profesores evalúan a los alumnos de forma individual o en pequeños grupos. De esta manera, desde una perspectiva de investigación, mientras estos diagnósticos pueden ser muy útiles para proporcionar una medida directa de las habilidades, hay que tener cuidado en su planificación, para tener la certeza de que se han asignado los recursos suficientes para proporcionar unos resultados fiables.

Diagnóstico a través de la carpeta. Los trabajos contenidos en la carpeta son una determinada recogida y evaluación de los trabajos de un estudiante que documentan el progreso a través de los objetivos de la enseñanza. Las carpetas se han empleado durante años en campos como, por ejemplo, la arquitectura, el arte y el periodismo, como el principal método de aprendizaje y logro. En la educación, el análisis de las carpetas ha ido aumentando, especialmente para el diagnóstico de las habilidades de lectura y escritura.

Debido a que las carpetas tienen la ventaja de proporcionar muchas pruebas del trabajo del estudiante a lo largo del tiempo, lo que puede usarse para evaluar su progreso, desde una perspectiva psicométrica su fiabilidad a menudo es baja. La puntuación de las carpetas se hace subjetivamente de acuerdo con la puntuación de las líneas guía o epígrafes, y es muy difícil obtener una fiabilidad inter-ratio alta. Esto puede deberse a la puntuación de criterios que son demasiado generales, incluso en programas a nivel nacional. Si los profesores puntúan localmente las carpetas desarrolladas, incluso pueden esperarse más errores de puntuación.

Finalmente, la evidencia para la validez tiene que ser considerada con cuidado ya que, normalmente, existe un deseo de generalizar características y objetivos extensos de la educación a partir de casos concretos. Por ejemplo, si los juicios se hacen sobre la capacidad de los estudiantes para expresarse por escrito, y los únicos tipos de escritos que hay en la carpeta son creativos y de exposición, la

validez de la inferencia sobre su expresión escrita más general es débil (constructo infraestimado).

INVENTARIOS DE PERSONALIDAD, ACTITUD, VALOR E INTERÉS

Los tests de aptitud y rendimiento son tipos de medición cognitiva. Los instrumentos afectivos o no cognitivos miden rasgos como, por ejemplo, intereses, actitudes, autoestima, valores, personalidad y opiniones. La mayoría está de acuerdo en que estos rasgos son importantes para el éxito escolar, pero medirlos de forma exacta es más difícil que la evaluación de rasgos o destrezas cognitivas. Primero, los resultados de los test no cognitivos pueden estar afectados de forma negativa por un conjunto de respuestas, que es la tendencia de la respuesta de un sujeto a ser influenciada por un «conjunto» general mientras responde a los ítems. Hay varios tipos de conjuntos de respuesta, incluyendo respuestas positivas o negativas a todas las preguntas, independientemente del contenido de los ítems; adivinación; y sacrificar la velocidad en favor de la exactitud. Un conjunto de respuestas es particularmente predominante con ítems ambiguos o ítems que usan de forma continua el acuerdo-desacuerdo o favorable-desfavorable. Segundo, los ítems no cognitivos son susceptibles de falsificarse. Aunque haya algunas técnicas que ayudan a reducir la falsificación, como por ejemplo, el empleo de preguntas de elección obligada, la ocultación de la finalidad del test y establecer un buen entendimiento con los sujetos, la falsificación siempre es posible. Uno de los tipos más serios de falsificación es la **deseabilidad social**, por la que los sujetos responden a los ítems para parecer más normales o más deseables socialmente en lugar de responder honestamente. Tercero, la fiabilidad de los tests no cognitivos, generalmente, es menor que la de los test cognitivos. Cuarto, en la mayoría de los test no cognitivos, nuestro interés se centra en la evidencia de la validez del constructo, lo que es difícil de establecer. Finalmente, los tests no cognitivos no tienen respuestas «correctas» como sucede en los tests cognitivos. Los resultados se interpretan, normalmente, por comparación con otros individuos, por lo que resulta particularmente importante la naturaleza del grupo de comparación. A pesar de estas limitaciones, los rasgos no cognitivos se utilizan en la investigación porque son una parte integrante del proceso de aprendizaje.

Los test de personalidad incluyen un amplio rango de listas de comprobación, tests descriptivos e inventarios de ajuste general. La mayoría son instrumentos de autopresentación que contienen un formato estructurado de pregunta-respuesta, y que requieren una preparación especial para su interpretación. Debido a la debilidad psicométrica de la mayoría de los test de personalidad, los resultados deben utilizarse para grupos de sujetos en vez de para individuos. Ver la tabla 6.6 para algunos ejemplos de test de personalidad.

Los inventarios de actitud y de interés se emplean de forma considerable en investigación educativa. La mayoría son instrumentos de autopresentación y están sujetos a falsificación y a un conjunto de respuesta. Los inventarios de interés miden sensaciones y opiniones sobre actividades en las que un individuo puede involucrarse. Los inventarios de actitud miden sentimientos y opiniones

TABLA 6.6: EJEMPLOS DE PRUEBAS NO COGNITIVAS

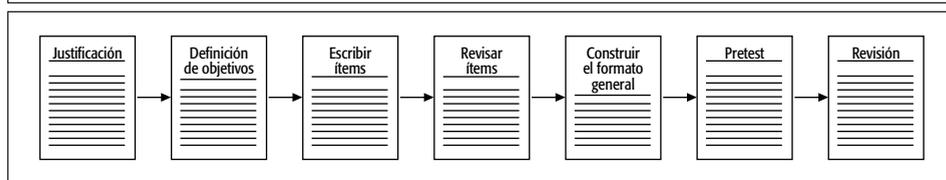
Personalidad	Actitud	Valores	Interés
Inventario de ajuste	Encuesta sobre los hábitos de estudio y actitudes	Estudio de valores	Inventario de intereses <i>Strong-Campbell</i>
Inventario multifásico de personalidad de Minesota	Encuesta sobre actitudes escolares	Encuesta de valores <i>Rokeach</i>	Inventario de intereses vocacional de Minesota
Inventario psicológico de California	Evaluación de la afectividad en la escuela de Minesota	Encuesta de valores de <i>Gordon</i>	Encuesta de interés ocupacional de <i>Kuder</i>
Inventario de personalidad para niños	Escala infantil de sociabilidad	Inventario de los valores del trabajo	Inventario de interés general de Kuder
Inventario de personalidad Ómnibus	Inventario sobre el conocimiento ambiental		Inventario de preferencia vocacional
<i>Test Inkblot Rorschach</i>	Inventario sobre actitudes de los escolares		
Test de percepción temática	Escala de actitud ante las matemáticas revisadas		
Escala de autoestima de Tennessee			
Escala de autoestima para niños de <i>Piers-Harris</i>			
Inventario de autoestima de <i>Coopersmith</i>			

sobre algo distante de una actividad, como por ejemplo, un objeto, un grupo o un lugar. Ambos se interesan por preferencias y por predisposiciones. Más allá del ámbito de este libro existe un debate completo sobre estos tipos de inventarios, aunque discutamos los cuestionarios como una forma de asegurar actitudes. La tabla 6.6 enumera ejemplos de inventarios de actitud, interés y valor.

CUESTIONARIOS

La técnica más ampliamente utilizada para obtener información de los sujetos es el cuestionario por muchas buenas razones. Un cuestionario es relativamente económico, incluye las mismas preguntas para todos los sujetos y puede asegurar el anonimato. Los cuestionarios manejan preguntas o enunciados, pero en todos los casos, el sujeto responde a algo escrito para un propósito concreto. En esta sección, presentamos información sobre los cuestionarios exponiendo la secuencia de pasos que siguen los investigadores cuando los desarrollan. Los pasos se resumen en la figura 6.1.

FIGURA 6.1: PASOS PARA EL DESARROLLO DE UN CUESTIONARIO



Justificación

Un cuestionario es una de las muchas formas con las que se puede obtener información. El investigador que quiere emplearlo, debe asegurarse que, dadas las limitaciones de la situación, no hay otras técnicas más fiables que puedan utilizarse. Esta decisión se basa en el conocimiento de los puntos fuertes y debilidades de cada técnica (que se trata más tarde en el capítulo por comparación con las técnicas más comunes). Los investigadores deberían pensar mucho para justificar siempre que desarrollan un nuevo cuestionario. En muchos casos podrían utilizarse instrumentos existentes o adaptarlos para usarlos en lugar de preparar nuevos. Si el investigador puede localizar un cuestionario existente, ahorrará mucho tiempo y dinero y podrá localizar un instrumento con validez y fiabilidad establecidas.

Definición de objetivos

El segundo paso para el empleo de un cuestionario es definir y enumerar los objetivos específicos que conseguirá con la información. Los objetivos se basan

en los problemas o cuestiones de la investigación, y muestran como se usará cada unidad de información. Deben ser objetivos conductistas estrictos, pero deben ser suficientemente específicos para indicar como cada ítem cumple los objetivos. Con la definición de los objetivos, el investigador especifica la información que necesita. Desafortunadamente, muchos investigadores incluyen cuestiones que no han sido meditadas lo suficiente y cuyos resultados no usarán nunca. Es un esfuerzo inútil, y provoca el desengaño de las audiencias interesadas.

Redacción de cuestiones y enunciados

Una vez que el investigador ha definido los objetivos y ha averiguado que no existen instrumentos que puedan usarse, comienza la tarea de redactar las cuestiones o enunciados. Lo mejor es redactar los ítems por objetivo y asegurarse de la forma en que se analizarán los resultados una vez que se hayan recogido los datos. Hay dos consideraciones generales en la redacción de los ítems: cumplir las reglas para la redacción de la mayoría de los ítems, y decidir qué formato de ítem es el mejor.

Babbie (1998) sugiere las siguientes líneas de orientación para la redacción de cuestiones o enunciados efectivos:

1. **Elaborar ítems claros.** Un ítem logra claridad cuando todos los entrevistados lo interpretan de la misma manera. Nunca asumir que el entrevistado leerá algo en el ítem. A menudo, las perspectivas, palabras o frases que tienen sentido para el investigador no son claras para los entrevistados. También puede ser que el ítem sea demasiado general, permitiendo diferentes interpretaciones. Por ejemplo, la pregunta, «¿qué piensas sobre el nuevo currículo?» puede evocar innumerables preguntas: por ejemplo, «¿qué currículo?», «¿qué significa 'pensar sobre'?»». Finalmente, las palabras imprecisas y ambiguas como, un poco, algunas veces, normalmente, deben evitarse, como también el lenguaje técnico y las frases complejas.

2. **Evitar las preguntas con dos objetivos.** Una pregunta debe limitarse a una sola idea o concepto. Las **preguntas con dos objetivos** contienen dos o más ideas, y frecuentemente se usa la palabra *y* en el ítem. Las preguntas y enunciados con dos objetivos no son deseables porque los entrevistados pueden, si se les da la oportunidad, responder cada parte de forma diferente. Si, por ejemplo, se le pide a un entrevistado que diga si está de acuerdo o no con el enunciado: «los asesores escolares gastan demasiado tiempo en la gestión de documentación administrativa y no el suficiente en asesorar sobre los problemas personales» puede estar de acuerdo con la primera parte (demasiada documentación administrativa) y en desacuerdo con la segunda idea (no demasiado tiempo asesorando).

3. **Los entrevistados deben ser competentes para responder.** Es importante que los entrevistados sean capaces de proporcionar información fiable. Algunas

preguntas efectuadas por los profesores para recordar acontecimientos determinados o reconstruir lo que hicieron varias semanas antes, por ejemplo, los sujetos son inexactos simplemente porque los profesores no pueden recordar de forma fiable los acontecimientos. De forma similar, sería de poco valor pedir a los profesores de la universidad que enseñan los fundamentos históricos de la educación que juzgaran la suficiencia de una prueba mínima de aptitud de habilidades de lectura para que maestros prospectivos demostraran conocimiento para la certificación. En muchos ejemplos, los sujetos no pueden dar una respuesta de la que estén convencidos, en cada una de las circunstancias lo mejor es proporcionar en las opciones de respuesta algo parecido a no estoy seguro o no lo sé para dar a los sujetos la oportunidad de expresar sus verdaderas sensaciones u opiniones.

4. **Las preguntas deben ser pertinentes.** Si se les pide a los sujetos que respondan a preguntas que no son importantes para ellos o que tratan sobre asuntos sobre los que no han pensado, lo más probable es que contesten de forma descuidada, y los resultados serán erróneos. Esto puede ocurrir, por ejemplo, cuando se les pregunta a los profesores por sus preferencias en cuanto a tests estandarizados, cuando rara vez emplean los resultados de estos test en la enseñanza. Sus respuestas pueden basarse en una respuesta de expediente en lugar de en una consideración cuidadosa de los tests.

5. **Los mejores ítems son los cortos y sencillos.** Los ítems largos y complicados deben evitarse debido a que son más difíciles de entender, y los entrevistados podrían ser remisos a intentar entenderlos. Hay que asumir que los entrevistados leerán y responderán los ítems rápidamente, y es necesario redactar ítems que sean sencillos, fáciles de entender y fáciles de responder.

6. **Evitar los ítems negativos.** Los ítems negativos deben evitarse porque son fáciles de malinterpretar. Los sujetos podrían de forma inconsciente omitir la partícula negativa, por lo tanto, sus respuestas serán lo contrario de lo que se pretendía. Si los investigadores utilizan ítems negativos, deberían subrayar, poner en mayúsculas o en negrita la partícula negativa (no o **NO**).

7. **Evitar ítems o términos sesgados.** La forma en la que se redactan los ítems, o la inclusión de ciertos términos, puede favorecer unas respuestas en lugar de otras. Tales ítems se denominan sesgados y, por supuesto, deben evitarse. Hay muchas formas de sesgar un ítem. La identificación de una persona o agencia muy conocida en el ítem puede crear sesgo. «¿Está usted de acuerdo o en desacuerdo con la reciente propuesta del inspector para...?», probablemente suscita una respuesta basada en la actitud hacia el inspector, no hacia la propuesta. Algunos ítems proporcionan respuestas sesgadas debido a la conveniencia social de la respuesta. Por ejemplo, si pregunta a los profesores si alguna vez han ridiculizado a sus alumnos, puede estar bastante seguro que, incluso aunque la respuesta fuese anónima, será *no* debido a que los buenos profesores no ridiculizan

a sus alumnos. Las respuestas de los alumnos a la misma cuestión o las observaciones de otros profesores podrían proporcionar diferente información.

Los investigadores también pueden proporcionar un indicio de la respuesta que esperan. Esto ocurre si los entrevistados quieren hacerle un favor al investigador y proporcionarle las respuestas que creen que espera, o puede ocurrir si los sujetos conocen las consecuencias de las respuestas. Se ha demostrado, por ejemplo, que las evaluaciones de los profesores realizadas por los alumnos son más favorables si el profesor les dice a los alumnos antes de que rellenen los formularios que los resultados influirán directamente en su (de los profesores) contrato y en su salario. Los alumnos se sentirán presumiblemente menos negativos debido a las importantes consecuencias de los resultados. Finalmente, los ítems son ambiguos si el entrevistado piensa: «bien, algunas veces me siento de esta manera, algunas veces me siento de aquella manera», o «depende de la situación». Muchos ítems fallan al especificar de forma adecuada las limitaciones situacionales que deben considerarse, y conducen a respuestas inexactas. Por ejemplo, si preguntamos si se está de acuerdo o no con el enunciado «el método de descubrimiento en la enseñanza es mejor que el método de lectura», un profesor podría responder: «depende del alumno».

Una vez expuestas las líneas de orientación generales, ¿cómo sabe si los ítems están bien redactados? Una aproximación es solicitar a los amigos, colegas y expertos que revisen los ítems y busquen cualquier problema. Por encima de este método subjetivo, una buena manera de demostrar empíricamente que los ítems son claros y no tienen ni sesgo, ni ambigüedad es construir dos formularios equivalentes de cada ítem y dar los ítems a grupos aleatorios. Si las respuestas de los dos grupos son casi las mismas sobre cada par de ítems, entonces los ítems probablemente son buenos. Si no, entonces los ítems tienen que redactarse de nuevo.

Tipos de ítems

Hay muchas formas de posible redacción de una pregunta o enunciado y varias formas de contestarla. El tipo de ítem debe basarse en las ventajas, empleos y limitaciones de estas opciones. A continuación, presentamos un resumen de las orientaciones más frecuentes sobre cómo deben redactarse y contestarse las preguntas y los enunciados.

Formato abierto y cerrado. La primera consideración es decidir si el ítem tendrá un **formato cerrado**, en el que el sujeto elige entre respuestas predeterminadas, o un **formato abierto**, en el que los sujetos escriben la respuesta que quieren. La elección del formato a usar depende del objetivo del ítem y de las ventajas o desventajas de cada tipo. Los ítems de formato cerrado (también denominados estructurados, selección de respuesta o cerrado-finalizado) son los mejores para la obtención de información demográfica y datos que pueden clasificarse fácilmente.

Por ejemplo, en lugar de preguntar, «¿cuántas horas estudió para el test?». Una pregunta con formato cerrado proporcionaría rangos de horas y pediría al entrevistado que marcara la casilla correspondiente, como se indica a continuación:

Marque la casilla que indica el número de horas que empleó estudiando para el test.

- 0-2 horas
- 3-5 horas
- 6-8 horas
- 9-11 horas
- más de 12+

Obviamente, será mucho más sencillo puntuar un ítem de formato cerrado, y el sujeto podrá contestar los ítems más rápidamente. Por lo tanto, lo mejor es utilizar ítems de formato cerrado con un gran número de sujetos o un gran número de ítems. El investigador emplearía demasiado tiempo en clasificar varios cientos de respuestas abiertas, por no mencionar la subjetividad involucrada.

Sin embargo, existen ciertas desventajas en el empleo de ítems estructurados. Por ejemplo, ante la pregunta «¿cuántas horas estudió para el test?», si cada uno de los sujetos selecciona la casilla con la etiqueta entre 3 y 5 horas, el investigador pierde exactitud y variabilidad (ningún margen de respuesta a través de todas las categorías de respuesta) en este factor. En otras palabras, si las categorías se crean para no permitir a los sujetos indicar exactamente sus sensaciones u opiniones, el ítem no es muy útil. Esto ocurre con algunos ítems de elección forzada. Otra desventaja es que un ítem estructurado da pistas al entrevistado de las posibles respuestas. Si, por ejemplo, preguntamos, «¿por qué hizo el test tan mal?», los alumnos, si se emplea un formato abierto, podrían listar dos o tres factores relevantes que pensaban que eran importantes. Sin embargo, un formato estructurado, podría listar veinticinco factores y el estudiante marcaría cada uno de los que fue importante (como por ejemplo: *no estudié lo suficientemente duro; estaba enfermo; no tuve suerte*); el alumno puede marcar factores que podría haber omitido en el modo abierto. Una aproximación al caso en el que tanto el formato abierto como el cerrado tienen ventajas es utilizar preguntas abiertas-cerradas, primero con un pequeño grupo de sujetos para generar factores destacados, y luego, utilizar ítems cerrados, basados en las respuestas abiertas-cerradas, con un grupo mayor. Los ítems abiertos-cerrados ejercen, al menos, algo de control sobre el entrevistado y pueden capturar diferencias de idiosincracia. Si el propósito de la investigación es generar respuestas individuales específicas, el mejor formato es el abierto; si el propósito es proporcionar respuestas de grupo más generales, el mejor formato es el cerrado.

Ítems de escalas valorativas. Una **escala** es una serie de gradaciones, niveles o valores que describen varios grados de algo. Las escalas se emplean en gran medida en los cuestionarios porque permiten valoraciones bastante exactas de

pareceres u opiniones. Esto se debe a que muchas de nuestras creencias y opiniones se piensan en términos de gradaciones. Creemos algo firmemente o fijamente, o quizá tenemos una opinión positiva o negativa de algo.

El formato general de los ítems de escala valorativa es una pregunta o enunciado seguido por una escala de respuestas posibles. Los sujetos marcan el lugar de la escala que mejor refleja sus creencias u opiniones sobre el enunciado. El ejemplo más ampliamente utilizado es la **escala Likert** (pronunciado Lick-ert). En la escala Likert, el tronco incluye un valor o dirección y el entrevistado indica acuerdo o desacuerdo con el enunciado. Los ítems tipo Likert emplean diferentes escalas de respuesta; el tronco puede ser tanto neutral como direccional. Lo siguiente es un ejemplo de una escala Likert:

La ciencia es muy importante:

Muy de acuerdo	De acuerdo	Ni de acuerdo ni en desacuerdo	En desacuerdo	Muy en desacuerdo
----------------	------------	--------------------------------	---------------	-------------------

Debe señalarse que mientras el formato acuerdo-desacuerdo se usa ampliamente, también puede ser de apariencia engañosa. Por ejemplo, podríamos estar en desacuerdo con el enunciado: «la Sra. Jones es una buena profesora» porque es una profesora excelente.

Las escalas tipo Likert proporcionan una gran flexibilidad debido a que los elementos que describen la escala pueden variar para fijarse en la naturaleza de la cuestión o del enunciado. Aquí tenemos unos ejemplos:

La ciencia es:

Extremadamente importante	Muy importante	Importante/ sin importancia	Un poco	Muy
---------------------------	----------------	--------------------------------	---------	-----

¿Cada cuánto se organiza bien tu profesor?

Siempre	La mayoría de las veces	Algunas veces	Raramente	Nunca
---------	-------------------------	---------------	-----------	-------

¿Cómo mediría la actuación de Cindy?

Muy pobre	Pobre	Aceptable	Buena	Excelente
-----------	-------	-----------	-------	-----------

Indique el grado en el que su actuación fue un éxito o un fracaso.

Éxito extremo	Éxito	OK	Fracaso	Fracaso extremo
---------------	-------	----	---------	-----------------

Indique cómo se siente tras su actuación.

Inmensamente orgulloso	Un poco orgulloso	Ni orgulloso ni avergonzado	Algo avergonzado	Inmensamente avergonzado
Muy feliz	Algo feliz	Ni feliz ni triste	Algo triste	Muy triste

Los investigadores se preguntan, en ocasiones, si debe incluirse la elección indiferente en una escala de Likert. Mientras se sigan incluyendo ambas formas, lo mejor es incluir una categoría intermedia. Si no se incluye la elección indiferente y el entrevistado se siente así, entonces lo estamos obligando a tomar una decisión incorrecta o a no responder nada. Una elección forzada puede llevar al entrevistado a una frustración, sin embargo, el argumento para eliminar la elección neutra adquiere importancia en circunstancias en las que los entrevistados tienen la tendencia a agrupar las evidencias en la categoría central. Cuando se han redactado los ítems de un cuestionario, deben ponerse de forma aleatoria dentro de cada sección separada del mismo.

Una variación de la escala de Likert es la de **diferencial semántico**. Esta escala emplea pares de adjetivos, cada uno de los cuales es como un ancla en los extremos de una sola línea. En esta escala no se necesita una serie de descriptores; sólo se coloca una frase o palabra en cada extremo. La escala se utiliza para obtener reacciones descriptivas hacia un objeto o concepto. Es una escala que se construye fácilmente y puede completarse rápidamente por los entrevistados. Los ejemplos que siguen ilustran sus empleos característicos.

Matemáticas

Me gustan _____ No me gustan
 Difícil _____ Fácil

Mi profesor

Fácil _____ Difícil
 Injusto _____ Justo
 Entusiasta _____ Desmotivado
 Aburrido _____ No aburrido

El extracto 6.2 muestra cómo se emplea un diferencial semántico para evaluar respuestas de los alumnos ante varios aspectos de la escuela. Mediante esta prueba, el afecto se evalúa por la escala, anclado por los términos *agradable-desagradable*, mientras que el valor de la actividad se mide por *importante-sin*

importancia. En el extracto 6.3 se emplea otro tipo de diferencial semántico, uno que obviamente está orientado hacia los niños más pequeños. La escala está limitada pero proporciona un rango desde feliz hasta triste.

EXTRACTO 6.2

ESCALA DE DIFERENCIAL SEMÁNTICO

28 juegos famosos

importante ○ ○ ○ ○ ○ sin importancia
agradable ○ ○ ○ ○ ○ desagradable

29 arte

importante ○ ○ ○ ○ ○ sin importancia
agradable ○ ○ ○ ○ ○ desagradable

39 del gusto de otros alumnos

importante ○ ○ ○ ○ ○ sin importancia
agradable ○ ○ ○ ○ ○ desagradable

30 del gusto de los profesores

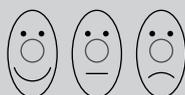
importante ○ ○ ○ ○ ○ sin importancia
agradable ○ ○ ○ ○ ○ desagradable

Fuente: *Minnesota School Attitude Survey* por el Dr. Andrew Ahlgren. Copyright © 1983 Science Research Associates. Reproducido con permiso de CTB Macmillan/McGraw-Hill.

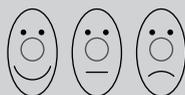
EXTRACTO 6.3

ESCALA DE DIFERENCIAL SEMÁNTICO PARA NIÑOS PEQUEÑOS

1. Enseñanza de aritmética



2. Enseñanza de ciencias



Fuente: From *Minnesota School Attitude Survey* por el Dr. Andrew Ahlgren. Copyright © 1983 Science Research Associates. Reproducido con permiso de CTB Macmillan/McGraw-Hill.

Ítems de ordenación. Un problema que puede aparecer con el empleo de la escala Likert o diferencial semántica es que todas las respuestas pueden ser la misma, haciendo más difícil diferenciar cada ítem. Si se emplea una escala Likert para investigar la importancia de cada una de las cinco maneras en que un

departamento de universidad gasta el dinero; un entrevistado, por ejemplo, podría marcar como *muy importante* para cada una. El resultado sería reducido para los esfuerzos del investigador en cuanto a cómo priorizar el gasto de dinero. Sin embargo, si se les pide a los entrevistados que clasifiquen las cinco formas por orden secuencial de mayor a menor importancia, entonces el investigador puede recoger información más valiosa sobre las formas de gastar el dinero. Una evaluación ordenada por rango del ejemplo anterior podría ser algo así:

Clasifique las siguientes actividades con respecto a su importancia en la asignación del presupuesto de investigación de este año. Use 1 = la más importante, 2 = siguiente más importante, y así sucesivamente hasta 5 = la menos importante.

- | | |
|--|---|
| <input type="checkbox"/> Coloquio anual | <input type="checkbox"/> Software informático |
| <input type="checkbox"/> Proyectos de investigación individuales | <input type="checkbox"/> Becas de estudiantes |
| <input type="checkbox"/> Ponentes invitados | |

Ítems de respuesta múltiple. Una lista de respuesta múltiple sólo es un método para proporcionar al entrevistado un número de opciones entre las que tiene que elegir. El ítem puede requerir la elección de una o de varias alternativas (por ejemplo: Marque una: El tema de biología que más me divierte es ecología, botánica, anatomía, microbiología, o genética.); o los entrevistados pueden marcar varias palabras como válidas:

Marque tantas como sean apropiadas. Los temas más divertidos en biología son:

- | | |
|---|--|
| <input type="checkbox"/> botánica | <input type="checkbox"/> ecología |
| <input type="checkbox"/> anatomía comparada | <input type="checkbox"/> microbiología |
| <input type="checkbox"/> genética | <input type="checkbox"/> zoología |

Las listas de respuesta múltiple también pueden utilizarse para pedir a los entrevistados que contesten si o no a una pregunta, o para marcar la categoría a la que pertenecen. Por ejemplo:

¿Esta casado/a? sí no

Marque la categoría apropiada:

- | | |
|--|-------------------------------------|
| <input type="checkbox"/> soltero | <input type="checkbox"/> separado |
| <input type="checkbox"/> no casado nunca | <input type="checkbox"/> divorciado |
| <input type="checkbox"/> casado | <input type="checkbox"/> viudo |

Observe que en las respuestas por categorías, los entrevistados sólo pueden situarse en una categoría y sólo en una.

Formato del ítem

Existen varias formas de presentar los ítems y las respuestas a los ítems. La presentación más clara es escribir el ítem en una línea y poner las categorías de respuesta debajo, no al lado del ítem. También es aconsejable emplear cajas de selección, llaves o paréntesis en lugar de una línea para indicar el lugar donde marcar. Por ejemplo:

¿Alguna vez has engañado?

- sí
- no

Es mejor que:

¿Alguna vez has engañado? ___ sí ___ no

Con las escalas de Likert y de diferencial semántico no se recomienda el empleo de líneas continuas o de espacios en blanco como zonas para marcar, debido a que la marca normalmente se sitúa entre dos opciones.

Algunas veces cuando un investigador plantea una serie de cuestiones, contestando una cuestión de una cierta manera se dirige a otras cuestiones. Eso se denomina **cuestiones de contingencia** y se ilustra a continuación:

¿Ha usado la Guía Curricular de Matemáticas?

- sí
- no

Si es que sí, ¿cuántas veces ha utilizado las actividades sugeridas?

- 0-2 veces
- 3-5 veces
- 6-10 veces
- más de 10 veces

¿Asistió a la Conferencia Nacional sobre Comprobación?

- sí (por favor, conteste preguntas 17-20)
- no (por favor, vaya a pregunta 21)

Si varias cuestiones emplearán el mismo formato de respuesta, como ocurre con los ítems de una escala Likert típica, es aconsejable construir una matriz de ítems y categorías de respuesta. En el extracto 6.4 se ilustra un ejemplo de esta matriz.

Formato general

La distribución y organización general del cuestionario es muy importante. Si parecen realizados sin cuidado o de forma confusa, los entrevistados probablemente lo apartarán y no lo contestarán nunca. Un formato o apariencia bien elaborado proporciona una primera impresión favorable y tendrá como resultado respuestas concienzudas, serias y cooperativas. Las siguientes reglas deben seguirse con sumo cuidado:

1. Revisar con cuidado la gramática, deletreo, puntuación y otros detalles.
2. Asegurarse de que la impresión es clara y que resulta fácil de leer.
3. Plantear instrucciones cortas y fáciles de entender.

EXTRACTO 6.4

MATRIZ DE PREGUNTA

Para las preguntas 1-8 utilice la siguiente escala de respuesta:

1	2	3	4	5	6
Nada en absoluto	Muy poco	Algo	Bastante	Extensamente	Completamente

En qué grado la evaluación del primer semestre de los alumnos de su clase estuvo basada en:

1. ¿La inclusión en la determinación de porcentaje final corrige si los alumnos no pudieron completar una tarea?	1	2	3	4	5	6
2. ¿El comportamiento estudiantil desorganizado?	1	2	3	4	5	6
3. ¿El comportamiento correcto del estudiante?	1	2	3	4	5	6
4. ¿Las actitudes de los alumnos hacia el aprendizaje?	1	2	3	4	5	6
5. ¿La mejora del comportamiento desde el comienzo del semestre?	1	2	3	4	5	6
6. ¿El escaso esfuerzo del alumno por aprender?	1	2	3	4	5	6
7. ¿El elevado esfuerzo del alumno por aprender?	1	2	3	4	5	6
8. ¿El grado de esfuerzo de los alumnos con poca capacidad?	1	2	3	4	5	6

4. Evitar abarrotar el cuestionario intentando incluir muchos ítems en una página.
5. Evitar ítems abreviados.
6. Hacer el cuestionario lo más corto posible.
7. Proporcionar un espacio adecuado para contestar a las cuestiones de respuesta abierta.
8. Usar una secuencia lógica y agrupar los ítems relacionados.
9. Numerar las páginas y los ítems.
10. Utilizar ejemplos si los ítems pueden resultar difíciles de entender.

11. Situar los ítems importantes cerca del principio en los cuestionarios largos.
12. Asegurarse de que la manera de situar y secuenciar las preguntas puede afectar a las respuestas.
13. Presentar las escalas de respuesta en cada nueva página.

Estudio piloto

Resulta muy recomendable que los investigadores realicen un estudio piloto de sus cuestionarios antes de emplearlos en los estudios. Lo mejor es localizar una muestra de sujetos con características similares a los que se usarán en el estudio. Mientras que el tamaño de la muestra debería ser mayor de veinte, es mejor tener sólo diez sujetos que no tener estudio piloto. La administración del cuestionario debería ser sobre lo mismo que se ha empleado en el estudio y en el estudio piloto, y los entrevistados del estudio piloto deberían tener espacio para incluir comentarios sobre cada ítem y sobre el conjunto del cuestionario. El investigador desea conocer si lleva mucho tiempo completarlo, si las instrucciones e ítems son claros, etcétera. Si hay suficientes sujetos en el estudio piloto, puede calcularse una estimación de la fiabilidad, y podrá darse alguna indicación acerca de la suficiente variabilidad en las respuestas para investigar varias relaciones. De esta forma son necesarios dos pasos para tener retroalimentación sobre un cuestionario antes de usarlo en un estudio: una crítica informal de los ítems individuales sobre cómo se han preparado y un estudio piloto del cuestionario completo.

ESQUEMAS DE UNA ENTREVISTA

Las entrevistas en los estudios cuantitativos, esencialmente, son cuestionarios orales. Los principales pasos para realizar una entrevista son los mismos que para preparar un cuestionario: justificación, definición de objetivos, redacción de preguntas, decisión del formato general y del ítem, y examen preliminar. La diferencia elemental es que las entrevistas implican una interacción directa entre los individuos, que tiene tanto ventajas como desventajas si las comparamos con los cuestionarios. La técnica de la entrevista es flexible y adaptable. Puede usarse con muchos tipos diferentes de asuntos y de personas, como por ejemplo, aquéllos que son analfabetos o demasiado jóvenes para leer y escribir, y las respuestas pueden probarse, seguirse, esclarecerse y elaborarse para conseguir respuestas específicas exactas. En las entrevistas pueden observarse tanto los comportamientos verbales como los no verbales, y el entrevistador tiene la oportunidad de motivar al entrevistado. Las entrevistas tienen como resultado una tasa de respuesta mucho más alta que los cuestionarios, especialmente sobre temas que tratan de las cualidades

personales o las sensaciones negativas. Para la obtención de información puntual y menos personal, son preferibles los cuestionarios.

Las principales desventajas de la entrevista son su potencial para la subjetividad y el sesgo, su coste y su consumo de tiempo y la falta de anonimato. Dependiendo de la preparación y la experiencia del entrevistador, el entrevistado puede no estar cómodo en la entrevista y reacio a mostrar sus verdaderos sentimientos; el entrevistador puede preguntar cosas relevantes para dar soporte a un punto de vista determinado; o las percepciones del entrevistador sobre lo que se expresó pueden ser inexactas. Debido a que entrevistar es una labor intensiva, es costosa y consume mucho tiempo (con la posible excepción de las entrevistas telefónicas), normalmente, se traduce en una muestra con menos sujetos de los que podrían obtenerse con un cuestionario, y puesto que las entrevistas implican a una persona hablando con otra, no es posible el anonimato. Puede acentuarse la confidencialidad, pero siempre existe el potencial para falsear, o para ser menos franco y sincero, ya que los sujetos pueden creer que compartir cierta información no sería apropiado para sus intereses.

Para mitigar el potencial de sesgo, el entrevistador debería interpretarse como un punto medio neutral a través del que se intercambia la información. Si se logra esta meta, entonces, la presencia del entrevistador no tendrá efecto sobre las percepciones o respuestas del entrevistado. En otras palabras, si la entrevista se hace correctamente, no importa quién sea el entrevistador; cualquier número de entrevistadores diferentes lograrían los mismos resultados. Este aspecto de la entrevista es esencial para la fiabilidad. Si dos o más entrevistadores están de acuerdo sobre la forma en que deben clasificarse la mayoría de las repuestas, entonces el proceso es fiable, como se asegura en el acuerdo de inter-ratio. También se puede obtener una estimación de estabilidad de la fiabilidad correlacionando los resultados de una entrevista inicial con los resultados obtenidos en un segundo momento por el mismo entrevistador. Puede utilizarse otro planteamiento para incrementar la exactitud de las entrevistas permitiendo al entrevistado revisar las percepciones del entrevistador. Esto puede lograrse si el entrevistador escribe sus percepciones de la respuesta a cada pregunta y las envía a los entrevistados. Los entrevistados pueden leer las percepciones escritas y hacer correcciones o añadidos donde crean apropiado. Una ventaja adicional de este planteamiento es que ayuda a construir una relación positiva entre el entrevistador y el entrevistado, lo que resulta de gran ayuda si el entrevistador siguiese con las entrevistas iniciales o estuviese envuelto en un estudio o evaluación continuos.

Preparación de la entrevista

Una vez que el investigador toma la decisión de recurrir a una entrevista para la recogida de datos, se elabora un esquema de la entrevista. El esquema enu-

mera todas las cuestiones que se preguntarán, dejando espacio para que el entrevistador escriba las respuestas. Las preguntas se relacionan directamente con los objetivos del estudio y siguen una secuencia determinada adherida a cada entrevista. En la mayoría de los casos, las preguntas escritas son exactamente las mismas que se preguntarán de forma oral, con las preguntas de análisis apropiadas. Normalmente las preguntas son de uno de estos tres tipos: estructuradas, semiestructuradas o sin estructurar. Las **preguntas estructuradas** (denominadas también preguntas de *respuesta limitada* o de *respuesta seleccionada*) van seguidas de un conjunto de opciones, y el entrevistado selecciona una de las opciones como la respuesta (por ejemplo, «¿diría que el programa ha sido muy efectivo, algo efectivo o nada efectivo?»). Las **preguntas semiestructuradas** no tienen opciones entre las que el entrevistado elige la respuesta. Más bien, la pregunta se redacta para permitir respuestas individuales. Es una pregunta de respuesta abierta pero es medianamente específica en su intención (por ejemplo, «¿cuál ha sido el aspecto más beneficioso del programa de enseñanza de tu profesor?»). Las **preguntas no estructuradas** permiten al entrevistador una gran flexibilidad para plantear una amplia variedad de preguntas en el orden que considere más apropiado. En los estudios educativos cuantitativos, la mayoría de las entrevistas utilizan una combinación de preguntas estructuradas y semiestructuradas, lo que proporciona un alto grado de objetividad y uniformidad, permitiendo, incluso, la comprobación y aclaración.

Después de redactar las preguntas, se necesita un estudio piloto para comprobar el sesgo en los procedimientos, el entrevistador o las preguntas. Durante el estudio piloto los procedimientos deben ser idénticos a los que se implementarán en el estudio. El entrevistador debería tomar especial nota de cualquier indicio que sugiera que el entrevistado no está cómodo o no entiende completamente la pregunta. Después de la entrevista, el entrevistado puede evaluar las preguntas por intención, claridad, etcétera. El estudio piloto proporciona los medios para evaluar la longitud de la entrevista y dará alguna idea al investigador de la facilidad con la que pueden ser resumidos los datos.

Un problema potencial que puede tratarse antes de que comience el estudio es la eliminación o la reescritura de las preguntas principales. Una **pregunta principal** se redacta para que el entrevistado esté más seguro de una respuesta que de otra, o si contiene información que puede sesgar la respuesta (como lo que se resumió antes para los ítems del cuestionario). Por ejemplo, si el investigador pregunta: «Teniendo en cuenta el gasto de adoptar una nueva colección de lectura, ¿deberíamos hacerlo este año?», obviamente, la redacción hace deseable contestar *no*. O consideremos la pregunta: «¿Está a favor de la comida caliente en el colegio?». Probablemente obtendremos una respuesta afirmativa. Como en el caso de los cuestionarios, la mejor manera de evitar preguntas principales en una entrevista es solicitar retroalimentación a otros expertos y hacer un estudio piloto de las preguntas.

Una consideración final en la preparación de la entrevista es pensar de qué manera pueden influir en las repuestas las características personales del entrevistador. Sax (1979, p. 243-244) enumera las siguientes variables que han demostrado influir en las respuestas:

Variable	Efecto en la entrevista
1. Edad del entrevistador	<ul style="list-style-type: none"> a. El entendimiento mutuo es alto para entrevistadores jóvenes y entrevistados de mediana edad. La edad del entrevistado no importa para entrevistadores mayores. b. Con personas del mismo sexo hay una menor inhibición en la respuesta. c. Cuando las personas son de la misma edad pero de distinto sexo se da la mayor inhibición en la respuesta. d. Los entrevistadores entre veinticinco y cincuenta años hacen un trabajo mejor que los más jóvenes o los más viejos.
2. Diplomados	Los entrevistadores expertos en ciencias de la conducta son calificados como más exactos que los expertos en ciencias físicas; los menos calificados son los diplomados en arte, empresariales, derecho y humanidades.
3. Nivel educativo	Los entrevistadores licenciados son mejor calificados que los que no lo son, pero las diferencias son mínimas.
4. Experiencia en entrevistas	La exactitud de los entrevistadores aumenta conforme aumenta su experiencia.
5. Procedencia racial	Las respuestas de los negros difieren dependiendo de si son entrevistados por blancos o por otros negros.
6. Procedencia religiosa	Las respuestas negativas relacionadas con los judíos tienden a ser ocultadas si los entrevistadores se presentan con nombres judíos.
7. Sexo del entrevistador	Los hombres obtienen menos respuestas en una entrevista que las mujeres.
8. Nivel socioeconómico	Los entrevistadores de clase media presentan un mayor grado de conservadurismo entre los entrevistados de clase trabajadora que los entrevistadores de clase trabajadora.

Muchos estudios educativos emplean entrevistadores ingenuos o inexpertos. En esta situación no sólo las características personales del entrevistador proporcionarán un posible sesgo, excepto que exista un error potencial simplemente

porque el entrevistador no está cualificado para manejar entrevistas. Si se utilizan principiantes, lo mejor es proporcionar preparación y supervisión. Esto puede ser caro y consumir mucho tiempo pero incrementará la validez y fiabilidad del estudio. Para ver detalles sobre la preparación de los entrevistadores, ver Babbie (1998) y el *Manual del entrevistador* (1999).

Durante la entrevista

La apariencia es muy importante. Lo mejor para el entrevistador es vestir de acuerdo a las normas existentes o de una forma similar a los entrevistados, y de ninguna manera vestir para que el entrevistado pueda pensar que el entrevistador representa un punto de vista determinado. El entrevistador debe ser amigable, relajado y agradable, y debe parecer interesado en el bienestar de los entrevistados. Para dar respuestas honestas a las preguntas, el entrevistado debe sentirse cómodo con el entrevistador. Un comportamiento y una apariencia apropiada proporcionan una base para establecer una relación cómoda y de entendimiento mutuo. El entrevistador debería charlar unos pocos minutos para establecer una correcta relación.

El entrevistador debería explicar brevemente el propósito de la entrevista y preguntar si el entrevistado tiene cualquier tipo de pregunta o duda antes de plantear temas específicos. Luego, las preguntas se formulan al entrevistado con las mismas palabras que aparecen en el esquema de la entrevista. No deben formularse de otra forma.

Las preguntas deben leerse sin error ni tropiezos, de una manera natural sin forzar nada. Para conseguirlo, el entrevistador tiene que estar muy familiarizado con las preguntas y debería practicar leyéndolas en voz alta.

El entrevistador necesita reproducir las respuestas que el entrevistado da a las preguntas. La reproducción se hace normalmente en una de estas dos formas, por grabación de la voz en una cinta o por medio de notas escritas. Las respuestas grabadas pueden ser analizadas por varios jueces y pueden utilizarse para estimar la fiabilidad. La grabación de las respuestas en una cinta es, normalmente, la forma más útil para las preguntas de respuesta abierta. Una grabadora obviamente recogerá la información de forma más completa y objetiva que las notas, pero la mera presencia de una grabadora puede afectar a la entrevista y a las respuestas, especialmente si se hacen preguntas personales. Si las preguntas están muy estructuradas hay una menor necesidad de grabar las respuestas.

El método más frecuente utilizado para reproducir las respuestas es tomar notas basadas en las respuestas. Hay dos extremos con la toma de notas. En un extremo, el entrevistador puede intentar escribir la respuesta exacta que le han manifestado; y en el otro, el entrevistador espera hasta que el entrevistado acaba y luego reconstruye la respuesta a cada una de las preguntas. El problema de

tomar notas al pie de la letra es que lleva mucho tiempo durante la entrevista; por otro lado, la información se pierde cuando el entrevistador deja, exclusivamente, en manos de su memoria el escribir las respuestas después de la entrevista. La mayoría de los entrevistadores se sitúan en el punto medio de ambos extremos y durante la entrevista toman notas breves que amplían cuando se ha terminado la entrevista.

El **sondeo de las entrevistas** para fomentar la claridad de una respuesta es un nivel que, si se utiliza mal, puede llevar a respuestas incompletas o inexactas. El entrevistador debe dar el tiempo suficiente al entrevistado para que responda y debería evitar anticiparse o cortar una respuesta posible. El sondeo debe ser imparcial para que no afecte a la naturaleza de la respuesta.

Después de que hayan contestado a todas las preguntas, el entrevistador debería dar las gracias al entrevistado y darle tiempo para hacer comentarios o sugerencias relacionadas con el tema de las preguntas o con la entrevista en general. Ya que el entrevistado ha perdido tiempo para cooperar, es importante finalizar la entrevista de una manera positiva.

El extracto 6.5 está recogido de un estudio que utilizó entrevistas como técnica de recogida de datos. El extracto fue el material que se incluyó en la sección de métodos del artículo bajo el subtítulo «Entrevistadores y procedimiento». El propósito del estudio era asegurar las explicaciones de las causas que usan los niños para explicar el éxito o el fracaso. Los sujetos fueron niños de primer, tercer y quinto curso, y la segunda frase del extracto muestra como se realizaron las pruebas, además la sección de Procedimiento describe el entorno de las entrevistas. Las preguntas fueron semiestructuradas. A los niños se les contaba la historia y entonces se les preguntaba, «¿por qué obtuviste una calificación alta?» o «¿por qué realizaste un trabajo tan insuficiente?»

ESQUEMAS DE OBSERVACIÓN

En un sentido, todas las técnicas de recogida de datos implican algún tipo de observación. Luego, por término general, la palabra *observación* se emplea para describir los datos que se recogen, independientemente de la técnica utilizada en el estudio. Sin embargo, los métodos de investigación observacionales también se refieren a un método de recogida de información más específico que es muy diferente a las entrevistas o los cuestionarios. Al igual que en la técnica para la recogida de información, el método observacional descansa sobre los elementos que el investigador ve y escucha, en lugar de descansar sobre las respuestas de los sujetos a las cuestiones o los enunciados.

El papel del observador en la mayor parte de la investigación cuantitativa es permanecer indiferente del grupo o del proceso y, de esta manera, actuar como un **observador experimentado**. Por ejemplo, un investigador puede querer estudiar el

EXTRACTO 6.5

ENTREVISTANDO

ENTREVISTADORES

Se emplearon cuatro entrevistadores para recoger los datos: dos hombres graduados en Psicología, una mujer graduada en Psicología y una mujer con una diplomatura en Ciencias psicológicas. Todos los entrevistadores fueron presentados a los niños y estuvieron un tiempo observando cada una de las seis clases de las que habían sido seleccionados los sujetos.

Los entrevistadores fueron entrenados en el manejo del Esquema de codificación de *Elig-Frieze* (Elig y Frieze, 1975) para que pudieran reconocer respuestas que eran de codificación dudosa sobre una o más de una de las dimensiones causales. Cuando la respuesta del niño no estaba clara para los propósitos de la codificación, el entrevistador indagaba para esclarecer el significado exacto de la respuesta del niño. La primera prueba siempre era una simple repetición de la respuesta del niño, con sus propias palabras. En la mayoría de los casos esto puede provocar que el niño elabore de nuevo la respuesta inicial. Sin embargo, si el niño no elabora o si la elaboración deja cuestiones inciertas, el entrevistador debería plantear un conjunto de preguntas preparadas de antemano. Por ejemplo, la prueba más común utilizada para el esclarecimiento de la dimensión de estabilidad, con una respuesta como por ejemplo, «porque lo intentó». Si la prueba de repetición no resultaba eficaz, el entrevistador podía preguntar, «¿lo intentó siempre o sólo algunas veces?». Esto siempre lleva a una clarificación. Pero en general, se necesitaron pruebas iniciales en menos del 5% de las respuestas. Si durante la codificación de las entrevistas transcritas los jueces creían que la prueba del entrevistador había sesgado la respuesta del niño, estaban preparados para codificar sólo aquella parte de la respuesta que precedía a la intervención que la sesgó.

PROCEDIMIENTO

Los esquemas de entrevista se determinaron de antemano para incorporarlos a las consideraciones previas de la discusión del diseño. Los preparativos se hicieron con la administración del colegio para entrevistar a los niños en aulas vacías cerca de sus clases y para grabar todas las entrevistas. Cada niño fue entrevistado de forma individual. Los entrevistadores acompañaban a los niños desde su clase a la sala de la entrevista. Después de un periodo para hacer sentir al niño lo más cómodo posible preguntándole acerca de su familia, sobre lo que había escuchado sobre el experimento y si permitía que grabásemos su voz en una grabadora, comenzaba la entrevista. Todas las entrevistas comenzaban con una introducción por parte del entrevistador.

Voy a mostrarte dibujos que pueden utilizarse en los libros de cuentos para ayudar a contar las historias. Las historias serán sobre niños como tú. Y las historias serán sobre cosas que probablemente te han pasado a ti o a alguien que tú conoces. Lo que haré será mostrarte los dibujos y contarte un pequeño frag-

mento de la historia. Como conoces a muchos niños, me gustaría que me ayudes a hacer las historias lo más reales posible.

Usando un orden predeterminado al azar, se presentaron al niño las cuatro historias con resultados similares de una en una. Antes de contarle la historia al niño, el entrevistador mostraba en la mano la fotografía que representaba la historia. Cuando los entrevistadores pensaban que la respuesta del niño era de codificabilidad cuestionable, lo probaban. Cuando los entrevistadores estaban satisfechos y habían entendido la respuesta del niño, mostraban al niño la siguiente fotografía y el proceso continuaba. Cuando se completaban las cuatro historias, los entrevistadores acompañaban de nuevo a los niños a su case.

Fuente: Frieze, I. H., y Snyder, H. M. (1980), «Children's beliefs about the causes of success and failure in school settings», *Journal of Educational Psychology*, 72, 186–196. Copyright © 1980 por American Psychological Association. Reproducido con permiso del editor y de los autores.

ajuste del estudiante universitario de primer año a la vida del campus, observando el comportamiento de los estudiantes universitarios de primer año en varios ambientes como un desconocido, sin participar, simplemente recogiendo información.

El papel del observador también depende del grado de inferencia o juicio que se requiere. En un extremo, el observador hace observaciones de **inferencia alta**, lo que requiere que el observador realice juicios o deducciones basadas en los comportamientos observados. Lo que se registra con una observación de inferencia alta es el juicio del observador. Por ejemplo, una observación de inferencia alta de un profesor sería una valoración realizada por el director sobre factores como el manejo de la clase y el entusiasmo. El director observaría la clase y haría una valoración de excelente, buena, regular o pobre de cada uno de los dos aspectos. Por otro lado, una observación de **inferencia baja** requiere que el observador registre un comportamiento específico en un sentido más global, sin hacer juicios. De esta forma, el director podría registrar el número de amonestaciones o anotaciones que utiliza el profesor, como información manejada para juzgar su desenvolvimiento en el aula. Una observación de inferencia baja normalmente resulta más fiable, pero muchos argumentarían que es necesario elaborar juicios basados en la complejidad y en multitud de variables dentro de una clase, para que resulten observaciones válidas. Y entre medias, el papel del observador es realizar juicios (inferencia alta) y registrar los comportamientos específicos y el contexto que lleva a la inferencia implicada en el juicio.

Justificación

Las principales ventajas del empleo de los métodos observacionales son que el investigador no necesita preocuparse sobre las limitaciones del sesgo de la

auto-presentación, conveniencia social o conjunto de las respuestas, y la información no está limitada a lo que puede recuperarse exactamente de los sujetos. El comportamiento puede registrarse si se produce de forma natural. Esta ventaja es muy importante para la investigación diseñada para estudiar lo que ocurre en la vida real como oposición a los ambientes muy artificiales. Sin embargo, la investigación observacional es cara y difícil de dirigir de una forma fiable para un comportamiento complejo. Es relativamente fácil y sencillo registrar, de forma objetiva, un comportamiento sencillo, pero la mayoría de los estudios se enfoca hacia un comportamiento más complejo que resulta difícil de definir y de evaluar mediante observación. También surge el problema de la forma en que el observador afecta al comportamiento de los sujetos con su presencia en el ambiente.

Definición de unidades de observación

El primer paso para el desarrollo de un estudio observacional es definir, con términos precisos, lo que se observará. Empezando por la cuestión o el tema de investigación, se averiguan las variables que necesitan observarse. Si la cuestión o el problema son generales, como por ejemplo, «¿cuánto tiempo están los estudiantes comprometidos académicamente?», entonces el investigador debe concretar el propósito para obtener unidades específicas y medibles que puedan observarse. Debido a que resulta imposible observar todo lo que ocurre, el investigador debe decidir sobre las variables o unidades de análisis que son más importantes y, luego, definir el comportamiento para que puedan registrarse de forma objetiva.

Registro de observaciones

Una vez que el investigador define el comportamiento a observar, se selecciona el procedimiento de grabación. Existen cinco tipos: grabación de duración, grabación de cuenta de frecuencia, grabación de intervalo, observación continua y muestra de tiempo.

Registro de duración. En el **registro de duración**, el observador indica la cantidad de tiempo que dura un tipo particular de comportamiento. A menudo, se emplea un cronómetro para medir la duración del comportamiento. De esta manera, el investigador simplemente busca un tipo de comportamiento (por ejemplo: estar fuera de su sitio, hablando con los demás estudiantes) y registra la cantidad de tiempo que dura este comportamiento dentro de un periodo de tiempo determinado.

Registro de frecuencia. El **registro de frecuencia** se usa cuando el observador está interesado únicamente en la frecuencia con la que tiene lugar el

comportamiento, no el tiempo que persiste. Generalmente, el observador posee una lista de varios tipos de comportamiento que se registrarán y lleva la cuenta para indicar cada cuánto ocurre cada uno. Obviamente, este tipo de registro es el mejor cuando la duración del comportamiento es corta (de uno a cinco segundos).

Registro de intervalo. En el **registro de intervalo** se observa a un sujeto durante un tiempo determinado (normalmente desde diez segundos hasta un minuto) y se registran sus comportamientos. El observador podría indicar que cada tipo de comportamiento ha ocurrido o no, o puede registrar cuántas veces ocurre dentro de cada intervalo.

Observación continua. En la **observación continua**, el observador proporciona una breve descripción del comportamiento del sujeto durante un periodo extenso. La descripción se redacta en orden temporal y el observador debe decidir que tipo de comportamiento es importante.

Muestra de tiempo. En la **muestra de tiempo**, el observador selecciona, a partir de un esquema fijo o aleatorio, los periodos de tiempo que se emplearán para observar determinados tipos de comportamiento. Este procedimiento se utiliza en conjunto con cada uno de los cuatro procedimientos mencionados anteriormente.

Si es posible, lo mejor es localizar esquemas observacionales existentes que estén estandarizados en algún grado. Apparently, se han desarrollado cientos de esquemas y, debido a que han sido probados con estudios piloto y se han usado en estudios previos, es más probable que demuestren mayor validez y fiabilidad que los nuevos esquemas.

Preparación de los observadores

La limitación más importante de una observación completa se concreta en la persona que registra lo que se ve y se oye, el observador. La dificultad descansa en la obtención de observaciones que sean objetivas, sin sesgo, y exactas en el sentido de que el observador evitó influir en el comportamiento de los sujetos. La objetividad del observador depende, de algún modo, de la especificidad del comportamiento. Esto es, un tipo de comportamiento descrito como «bromear con otros estudiantes» es mucho menos específico y más difícil de interpretar que acciones más objetivas y concretas como «levantar la mano» o «levantarse de la silla».

En la investigación observacional, el sesgo es un factor del grado en que las percepciones de idiosincracia y las interpretaciones del observador, influenciadas por experiencias previas, afectan al registro del comportamiento.

A pesar de que resulta casi imposible eliminar el sesgo, existe una necesidad de controlarlo. Una manera de controlar el sesgo es elegir cuidadosamente a los observadores. Indiscutiblemente, sería una mala idea elegir, como observador de los efectos de la auténtica valoración, a un defensor de ese tipo de educación; como sería injusto escoger a un adversario reconocido, ya que sus nociones preconcebidas podrían influenciar fácilmente sus observaciones. Un segundo método para el control del sesgo es trabajar con observadores preparados cuidadosamente, comparando sus observaciones con las de otros en cada ambiente durante el estudio. Mientras que los observadores están independientemente de acuerdo, hay menos oportunidad de que el sesgo sea un factor de confusión, lo que puede ocurrir si el observador es conocedor de las especificaciones del estudio. Por ejemplo, en un estudio de las diferencias entre los denominados buenos y malos profesores, si el observador conoce, antes de realizar las observaciones, qué profesores se supone que son buenos y malos, este saber, probablemente, sesgue las observaciones. De este modo, lo mejor para los observadores es tener el menor conocimiento posible del propósito del estudio. Su trabajo consiste en observar y registrar de una manera distante y objetiva.

MEDICIONES NO REACTIVAS

Los cuestionarios, las entrevistas y la observación directa son medidas intrusivas o reactivas, en el sentido de que los participantes se dan cuenta de que se les está observando. La principal dificultad relacionada con el conocimiento de los sujetos que participan es que su comportamiento puede verse afectado por este conocimiento. Un tipo de medición, en la que no se les solicita a los sujetos que hagan nada fuera de lo normal, se denomina no reactiva. Las **mediciones no reactivas o discretas** proporcionan datos que no están influenciados por el conocimiento de los sujetos participantes o por una alteración en el curso natural de los sucesos. Los principales tipos de medidas discretas incluyen las huellas físicas como por ejemplo, conocimientos adquiridos, libros, ordenadores, documentos, informes, cartas y observación en la que el sujeto ignora ser investigado. Ver Webb, Campbell, Schwartz y Sechrest (2000).

En cada una de las cinco principales técnicas de recogida de datos que hemos tratado en este capítulo coinciden tanto ventajas como desventajas que se resumen en la tabla 6.7. Los investigadores necesitan considerarlas para elegir los métodos de recogida de datos apropiados. ¿Qué técnica proporcionará la medida más válida y fiable para el rasgo? ¿Qué técnica es la más fiable? ¿Cómo pueden reducirse los posibles defectos?

TABLA 6.7: VENTAJAS Y DESVENTAJAS DE LAS TÉCNICAS DE RECOGIDA DE DATOS

Técnica	Ventajas	Desventajas
Test de lápiz y papel	Económico. Cuestiones estándar. Tests comerciales con altas cualidades técnicas. Tests objetivos fáciles de puntuar. Los tests estándar proporcionan procedimientos uniformes para todos los sujetos y puntuaciones estándar.	Las normas pueden no ser apropiadas. Los test estandarizados pueden ser demasiado amplios y generales. Pueden proporcionar un sentido falso de validez. Los desarrollados particularmente pueden tener debilidad técnica. Ansiedad. Restringido a los que saben leer.
Valoración alternativa	Proporciona medida holística directa de niveles. Estrechamente sincronizado con las instrucciones. Usa contextos más auténticos.	Valoraciones subjetivas con poca fiabilidad. Muestreo pobre o gran dominio de niveles. Costosa. Lenta.
Cuestionario	Económico. Puede ser anónimo. Cuestiones estándar y procedimientos uniformes. Normalmente fáciles de puntuar. Proporciona tiempo para que los sujetos piensen las respuestas.	Incapacidad de probar y esclarecer. Puntuación de elementos con respuesta abierta. Falsificación y deseabilidad social. Restringido a sujetos que pueden leer y escribir. Elementos ambiguos o sesgados. Conjunto de respuestas.
Entrevista	Flexible. Adaptable. Capacidad de probar y esclarecer. Capacidad de incluir comportamientos no verbales. Alto grado de respuesta. Se emplea con analfabetos.	Costosa. Lenta. Sesgo del entrevistador. No es anónima. Efectos del sujeto. Efecto de las características del entrevistador. Requiere preparación. Cuestiones destacadas.

(continúa)

(continuación)

Técnica	Ventajas	Desventajas
Observación	Captura el comportamiento natural. Reduce la deseabilidad social, el conjunto de respuestas y los efectos del sujeto. Relativamente discreto. Fiable para observaciones de baja inferencia.	Costosa. Lenta. Efecto del observador sobre los sujetos. Requiere preparación. Sesgo del observador. Fiabilidad difícil para comportamiento complejo y observaciones de alta inferencia. Imposibilidad de probar y esclarecer. Normalmente no anónima. Interpretación de las observaciones de alta inferencia.

RESUMEN

Este capítulo ha introducido varias técnicas que se utilizan frecuentemente en la recogida de datos descriptivos cuantitativos. Estas técnicas se emplean en investigación básica, aplicada y evaluativa, y pueden utilizarse en investigación experimental y no experimental. Los principales puntos del capítulo son los siguientes:

1. La evidencia para establecer conclusiones válidas a partir de las puntuaciones del test puede resultar apropiada para el empleo de los resultados.
2. La evidencia para la validez se basa en cuatro componentes: contenido, procesos de respuesta, estructura interna y relación con otras variables.
3. Existen cinco tipos principales de fiabilidad que se emplean para valorar la consistencia de las puntuaciones: estabilidad, equivalencia, estabilidad y equivalencia, consistencia interna y acuerdo.
4. Los test estandarizados proporcionan procedimientos uniformes para su administración y su puntuación.
5. Los resultados de los tests referidos a normas se basan en la comparación de una puntuación con las puntuaciones de un grupo de referencia.
6. Los resultados de los tests referidos a criterios comparan una puntuación con un modelo de ejecución establecido.
7. Los tests de aptitud predicen el comportamiento.
8. Los tests de rendimiento miden el aprendizaje previo.
9. Los diagnósticos alternativos, que incluyen los basados en la ejecución y en la carpeta de trabajos, proporcionan una medida de niveles directa.

10. Los instrumentos no cognitivos miden personalidad, actitudes, valores e intereses.
11. Los cuestionarios escritos son económicos, pueden asegurar el anonimato y permiten el empleo de preguntas estandarizadas.
12. Los cuestionarios ya existentes, probablemente, poseen mayor fiabilidad y validez que los que desarrolla el investigador.
13. Los ítems de un cuestionario suelen basarse en objetivos específicos y deben ser claros, relevantes, cortos y ordenados. Deben evitarse los ítems y términos sesgados.
14. Los ítems poseen formato abierto o cerrado, dependiendo de los objetivos y la naturaleza de la información deseada.
15. Los ítems escalados, como por ejemplo los de Likert y los diferenciales semánticos, emplean graduaciones de respuestas.
16. Los cuestionarios son económicos y pueden ser anónimos.
17. Los esquemas de la entrevista proporcionan flexibilidad y la capacidad para probar y aclarar las respuestas; registran tanto el comportamiento verbal como el no verbal. Proporcionan un alto índice de respuesta pero son costosos y más susceptibles al sesgo.
18. Las cuestiones de la entrevista son estructuradas, semiestructuradas o no estructuradas. Cada tipo tiene sus ventajas y desventajas.
19. Los procedimientos observacionales pueden registrar el comportamiento de forma natural y evitar algunas de las desventajas asociadas con los cuestionarios y las entrevistas.
20. Es difícil establecer y mantener la fiabilidad y la validez en la investigación observacional.
21. Las observaciones de baja inferencia resaltan el registro objetivo del comportamiento, mientras las observaciones de inferencia alta requieren juicios muy subjetivos por parte de los observadores.
22. Los procedimientos de registro en la observación directa incluyen: duración, frecuencia, intervalo, continuidad y muestra de tiempo.
23. Las medidas no reactivas pueden utilizarse para la recogida de datos de un suceso sin interrupción del acontecimiento natural del suceso.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La evidencia basada en el contenido es la más apropiada para...
 - a. los tests predictivos.
 - b. los tests de rendimiento.

- c. los tests de aptitud.
 - d. la b y la c son correctas.
2. Los tests de relaciones criterioales se diferencian de los tests con referencia a criterios en que...
- a. los tests con referencia a criterios utilizan grupos de normas mientras que los test de relaciones criterioales manejan criterios.
 - b. el diagnóstico criterial es un tipo de fiabilidad; los tests con referencia a criterios utilizan grupos de normas.
 - c. el test de relación criterial sirve para la validez; la referencia a criterios sirve para la fiabilidad.
 - d. los tests con referencia a criterios manejan criterios; el test de relación criterial se emplea para la validez.
3. Las estimaciones de consistencia interna de la fiabilidad se emplean mucho porque...
- a. sólo se necesita una forma de un test.
 - b. el coeficiente de fiabilidad normalmente es muy alto.
 - c. son las más fácil de calcular.
 - d. son más exactas que las estimaciones de estabilidad y equivalencia.
4. Los tests estandarizados se diferencian de los procedimientos elaborados particularmente en que los tests estandarizados...
- a. tienen tanto normas como soporte técnico.
 - b. tienen mayor flexibilidad.
 - c. son más válidos.
 - d. tienen procedimientos de administración y puntuación uniformes.
5. El propósito de un test de aptitud es...
- a. medir la inteligencia.
 - b. predecir el comportamiento.
 - c. medir los múltiples aspectos de las ventajas potenciales.
 - d. asegurar el rendimiento.
6. Los inventarios no cognitivos generalmente...
- a. pueden falsearse.
 - b. son altamente fiables.
 - c. son altamente válidos.
 - d. son socialmente deseables.
7. Elija la secuencia correcta para el desarrollo de un cuestionario.
- a. definición de objetivos, justificación, redacción de los ítems, examen preliminar.
 - b. examen preliminar, definición de objetivos, justificación, redacción de los ítems.
 - c. justificación, examen preliminar, definición de objetivos, redacción de los ítems.

- d. justificación, definición de objetivos, redacción de los ítems, examen preliminar.
8. Todo lo siguiente va asociado a la redacción de enunciados o de preguntas excepto...
 - a. evitar ítems sesgados.
 - b. evitar ítems redactados negativamente.
 - c. usar cuestiones de doble sentido.
 - d. redactar ítems claros.
 - e. redactar ítems cortos.
 9. La deseabilidad social es una forma de...
 - a. sesgo.
 - b. falta de fiabilidad.
 - c. limitación circunstancial.
 - d. la a y la c son correctas.
 10. El formato de un cuestionario debería...
 - a. emplear todo el espacio disponible de cada página.
 - b. utilizar ítems abreviados para ahorrar espacio.
 - c. situar en primer lugar los ítems importantes en los cuestionarios largos.
 - d. acortar al máximo posible el número de páginas.
 11. Las cuestiones o enunciados de formato cerrado son las mejores si se emplean para...
 - a. información clasificada fácilmente.
 - b. estudios piloto iniciales.
 - c. estudios a gran escala.
 - d. la a y la c son correctas.
 12. La escala *Likert* y los ítems de diferencial semántico son similares en que ambos...
 - a. utilizan adjetivos opuestos.
 - b. son fáciles de construir.
 - c. son difíciles de puntuar objetivamente.
 - d. son escalas.
 13. Los cuestionarios necesitan probarse de forma piloto debido a que...
 - a. el examen preliminar establece la fiabilidad.
 - b. puede advertirse una variación insuficiente en las respuestas.
 - c. pueden advertirse los análisis de datos de prueba.
 - d. la a y la b son correctas.
 14. Las cuestiones semiestructuradas en las entrevistas se redactan en un formato que permite...
 - a. una repuesta libre a una pregunta concreta.
 - b. una respuesta que fuerza a una elección en una pregunta concreta.
 - c. una respuesta libre a una pregunta amplia.

- d. una respuesta que fuerza a una elección en una pregunta amplia.
15. Lo mejor para los entrevistadores es adaptar la redacción específica de las preguntas a las características del entrevistado. ¿Verdadero o Falso?
- verdadero.
 - falso.
16. Los entrevistadores siempre deberían vestir de manera formal (por ejemplo, traje y corbata). ¿Verdadero o Falso?
- verdadero.
 - falso.
17. El procedimiento más frecuente para el registro de respuestas en una entrevista es...
- una grabadora.
 - una videocámara.
 - escribir las respuestas de forma textual.
 - escribir notas.
18. El observador experimentado...
- intenta llegar a ser un miembro del grupo de sujetos.
 - no provoca reacciones.
 - está distante del entorno.
 - está completamente informado de la naturaleza del estudio.
19. La observación es más objetiva cuando el observador no tiene que marcar las respuestas en un formato estándar fácil de interpretar. ¿Verdadero o Falso?
- verdadero.
 - falso.
20. Puede controlarse el sesgo del observador mediante todo lo siguiente excepto...
- la elección cuidadosa de los observadores.
 - mantener informado al observador de la naturaleza del estudio.
 - preparando cuidadosamente a los observadores.
 - empleando dos o más observadores para cada situación.
21. Otra palabra para las medidas discretas podría ser...
- sesgadas.
 - faltas de fiabilidad.
 - no reactivas.
 - no sesgadas.

Problemas de aplicación

- Para cada uno de los siguientes casos indicar si sería más apropiado un cuestionario, una entrevista o una técnica de observación y justificar su respuesta.

- a. Razones por las que 1.500 parejas creen que tienen problemas en su matrimonio.
 - b. Las actitudes de los alumnos de séptimo curso hacia los niños de integración.
 - c. Conocimiento de los padres del currículo del colegio.
 - d. Edad media y experiencia de los directores de colegio.
 - e. Consecuencias sobre el comportamiento agresivo de ver programas violentos en la televisión.
 - f. Percepciones de los estudiantes universitarios de la efectividad de los anuncios en el vestíbulo de la residencia.
 - g. Actitudes de los niños de preescolar hacia sus padres.
 - h. Actitudes de los profesores hacia la enseñanza basada en la competencia.
2. Indicar cuál es incorrecto para los siguiente ítems de cuestionario:
- a. ¿Qué piensa sobre la educación abierta?
 - b. Ordene los enunciados de mayor a menor importancia.
 - c. Los profesores de instituto necesitan más preparación en las formas de motivar a los estudiantes:

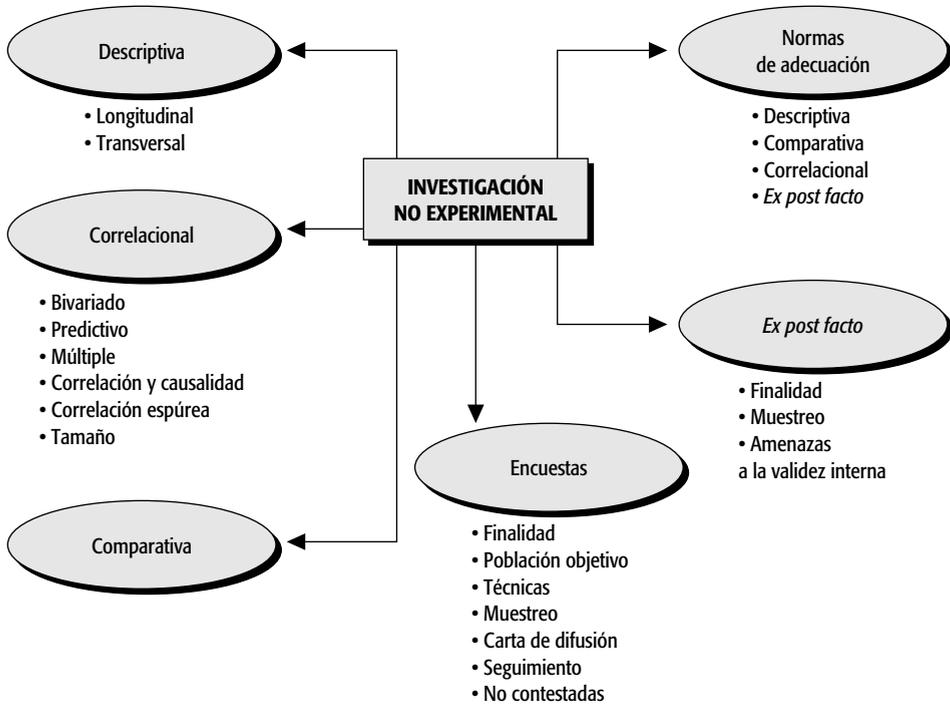
Totalmente de acuerdo	De acuerdo	En desacuerdo	Totalmente en desacuerdo
--------------------------	------------	---------------	-----------------------------

- d. El Sr. Jones es un buen profesor:

Totalmente de acuerdo	De acuerdo	En desacuerdo	Totalmente en desacuerdo
--------------------------	------------	---------------	-----------------------------

3. Para los profesores es importante observar y registrar las indicaciones que sus alumnos estudian e intentan aprender con el material asignado. Si un profesor de tercer curso se acerca y le pregunta cómo pueden hacerse tales observaciones con la mínima interrupción de la rutina diaria, ¿qué sugerencias le haría?

Diseños de investigación no experimental y encuestas



PALABRAS CLAVE

investigación descriptiva	coeficiente de regresión
estudios evolutivos	pesos beta
longitudinal	coeficiente de correlación múltiple
transversal	análisis de sendas
investigación comparativa	correlación espúrea
bivariada	atenuación
estudios de predicción	restricción en rango
variable predictora	coeficiente de determinación
variable criterio	investigación mediante encuesta
ecuación de predicción de regresión múltiple	investigación <i>ex post facto</i>

En el capítulo 2 se hacía una importante distinción entre modalidades de investigación experimental e investigación no experimental. En este capítulo analizaremos, minuciosamente, cuatro tipos elementales de diseños no experimentales, además de la investigación mediante encuesta.

INVESTIGACIÓN DESCRIPTIVA

La **investigación descriptiva** se ocupa de la situación actual o pasada de algo. Este tipo de investigación simplemente describe la realización, las actitudes, los comportamientos u otras características de un grupo de sujetos. Un estudio descriptivo pregunta *qué es o qué fue*; presenta las cosas de la manera *cómo son o eran*. La investigación descriptiva no implica el manejo de variables independientes.

La investigación descriptiva proporciona datos muy valiosos, particularmente, cuando se investiga un área por primera vez. Por ejemplo, se han realizado muchas investigaciones sobre la naturaleza del ambiente de la clase y su relación con las actitudes de los alumnos y el aprendizaje. Un primer paso de esta investigación fue describir de una forma adecuada lo que se entendía por «clima de la clase». Los estudios sobre el clima de la clase, que evalúan características como por ejemplo, el modo en que los alumnos hablan y actúan entre ellos, qué sienten hacia el profesor y sus sentimientos de franqueza, aceptación, confianza y respeto, se emplean para entender la atmósfera de la clase. Una vez se ha logrado esta comprensión descriptiva, pueden relacionarse varias dimensiones del clima con el aprendizaje de los estudiantes y la satisfacción del profesor y, por último, puede

controlarse el clima para examinar relaciones causales. Para muchos profesores, la sola descripción del clima puede resultar muy útil para entender la naturaleza de su clase. O suponga que quiere estudiar la relación entre los estilos de liderazgo de los directores y el estado de ánimo del profesor. De nuevo, el primer paso es describir los principales estilos de liderazgo. La pregunta descriptiva apropiada podría ser: ¿cuáles son los estilos de liderazgo de los directores? A continuación presentamos algunas cuestiones de investigación descriptiva:

- ¿Cuántos ejercicios realizan los alumnos?
- ¿Cuáles son las actitudes de los estudiantes hacia los niños de integración?
- ¿Con cuánta frecuencia engañan los estudiantes?
- ¿Qué piensan los profesores sobre la retribución por los méritos?
- ¿Cómo emplean su tiempo los alumnos durante el estudio independiente?
- ¿Cuáles son los componentes del programa para superdotados?

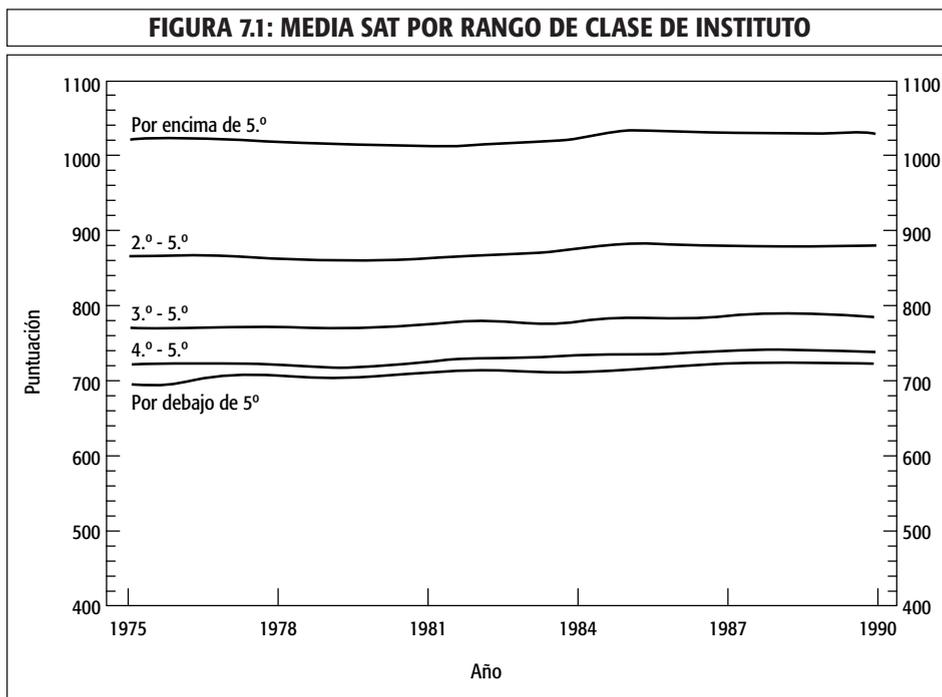
Cuestiones como éstas son muy importantes porque proporcionan la base para asuntos adicionales. Una vez que se ha definido un fenómeno adecuadamente, pueden tratarse cuestiones evolutivas, de diferencia y de relación. Los **estudios evolutivos** investigan los cambios de los sujetos a través del tiempo. El «mismo» grupo de sujetos puede estudiarse durante un espacio de tiempo (**longitudinal**) acerca de factores como variables cognitivas, socioemocionales o físicas. Por ejemplo, un estudio evolutivo longitudinal del desarrollo adulto comenzaría con la identificación de un grupo de adultos como sujetos; se medirían variables dependientes como intereses, metas, satisfacción, patrones de amistad, etc.; y después, se medirían estas variables en los mismos sujetos cada cinco años. En realidad, hay variantes en los diseños longitudinales, dependiendo de los sujetos que forman parte de la muestra o se utilizan para componer el «mismo» grupo. En un estudio de *tendencias* se estudia una población general a través del tiempo, aunque se toman muestras de la población cada año u otro intervalo temporal de recogida de datos. En un estudio longitudinal de *cohortes* se estudia a la misma población a lo largo del tiempo, mientras que en un estudio de *panel* los mismos individuos son encuestados cada vez que se recogen los datos. Los estudios evolutivos también pueden ser transversales, en los que diferentes grupos de sujetos (por ejemplo, en nuestro estudio de grupos adultos de 20, 30, 40 y 45 años) se estudian al mismo tiempo.

Una ventaja obvia de un estudio de panel es que, debido a que se examina el mismo grupo a través del tiempo, se asegura la posibilidad de comparación de los sujetos. Otra ventaja es que los sujetos responden a circunstancias, actitudes y opiniones presentes, en lugar de intentar recordar el pasado. Una desventaja de los estudios de tendencia es que la población de interés puede cambiar de un año al siguiente, lo que podría afectar a los resultados. Un buen ejemplo de esto es el análisis de tendencia de puntuaciones SAT. Aunque es cierto que la puntuación

media SAT para todos los estudiantes que han hecho el test ha descendido (hasta la reforma en 1995), este declive es más una función de quién hizo el test que del rendimiento de los estudiantes. Como se ilustra en las figuras 7.1 y 7.2, el declive está causado, realmente, por más estudiantes que eran menos capaces de hacer el test. De hecho, cuando se analizan las puntuaciones SAT de los resultados en los niveles altos de escolarización, las puntuaciones han permanecido relativamente estables; no muestran claramente un descenso.

Una gran desventaja de la investigación longitudinal es que lleva mucho tiempo completarla e implica compromisos significativos de tiempo, de dinero y de recursos del investigador. También es difícil conservar un registro de los sujetos y mantener su cooperación durante un largo periodo de tiempo. Los investigadores implicados en investigaciones transversales pueden estudiar grupos más numerosos con un coste menor, todo al mismo tiempo. De hecho, no necesitan esperar años para completar la investigación. La principal desventaja de la investigación transversal es que las diferencias en la selección entre los grupos pueden sesgar los resultados.

Una limitación importante de los estudios descriptivos es que las conclusiones relacionales (de relaciones entre variables) no están garantizadas. Es fácil

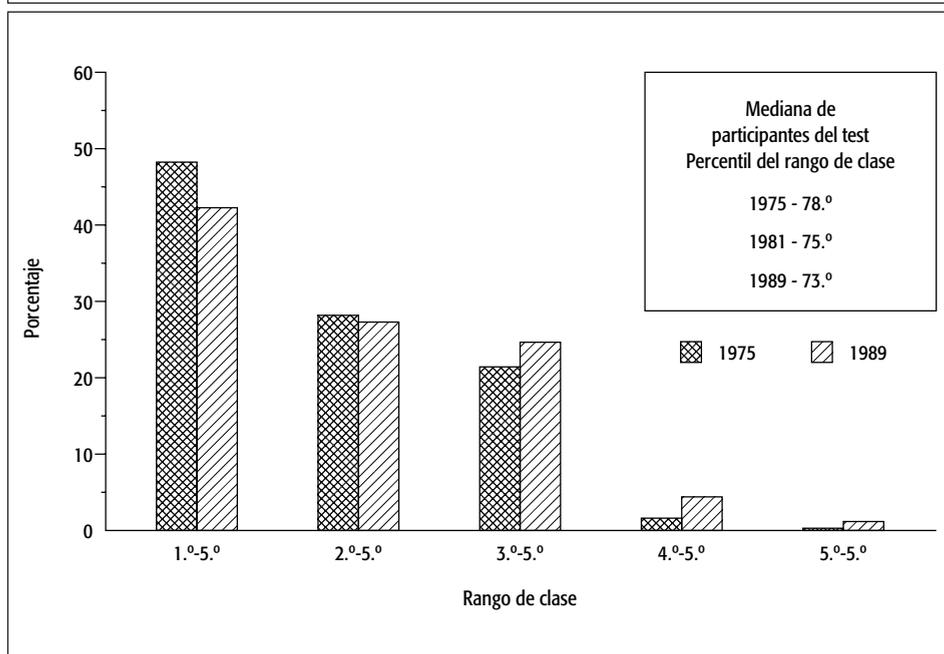


Fuente: *Perspectives on Education in America* (1991). Albuquerque, NM: Sandia National Laboratories.

hacer suposiciones a partir de descripciones simples sobre cómo se relacionan dos o más variables, pero esto llevará a conclusiones erróneas. Por ejemplo, suponga que un estudio describe los tipos de preguntas que profesores y alumnos plantean en clase y describe que los profesores principalmente hacen preguntas de «bajo nivel» y los estudiantes preguntan poco o nada. Sería tentador concluir que existe una relación entre estos fenómenos, esto es, una relación entre la edad o la posición y el tipo de preguntas realizadas. Sin embargo, uno no puede llegar a tal conclusión a menos que los estudiantes pregunten un cierto tipo de pregunta en lugar de no preguntar en absoluto.

En la investigación descriptiva, también, es importante prestar gran atención a la naturaleza de los sujetos y de las pruebas. Por ejemplo, debería saber si la muestra se realizó con voluntarios y si los resultados podrían haber sido diferentes si se hubiesen incluido otros sujetos. Las técnicas de recogida de datos necesitan buscar la fiabilidad y la validez, y deberían especificarse los procedimientos para la recogida de información. Es necesario especificar cuándo se recogieron los datos, por quién y bajo qué circunstancias. Por ejemplo, la descripción de una clase puede ser diferente dependiendo de si el observador es un profesor, un padre o un director.

FIGURA 7.2: PORCENTAJE DE ALUMNOS QUE HICIERON EL SAT POR RANGO DE CLASE

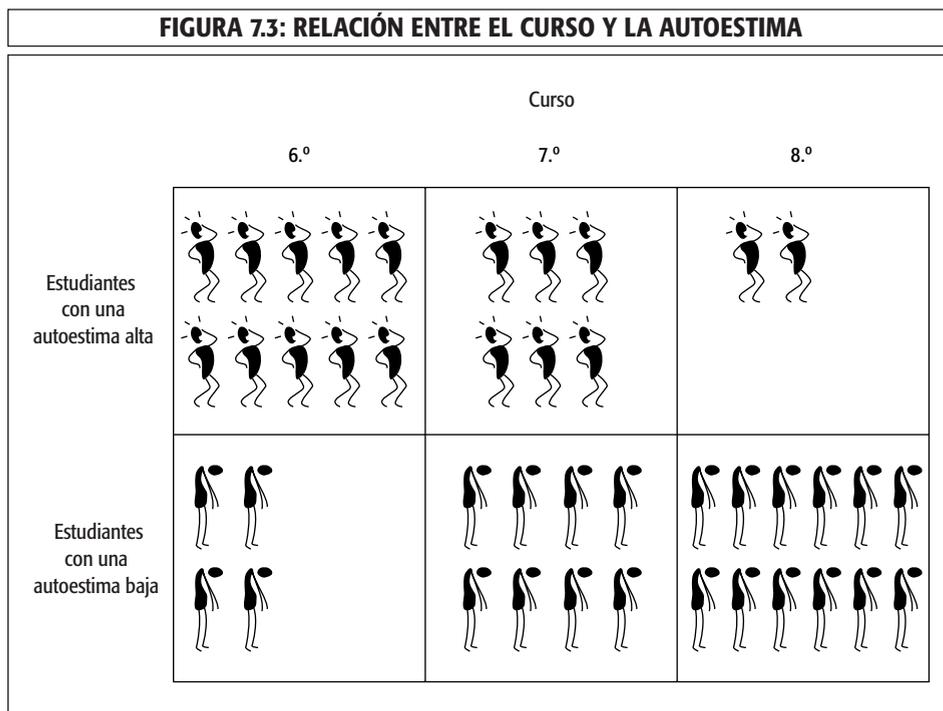


Fuente: *Perspectives on Education in America* (1991). Albuquerque, NM: Sandia National Laboratories.

RELACIONES EN LA INVESTIGACIÓN NO EXPERIMENTAL

Antes del examen de los estudios comparativos y correlacionales, debemos aclarar la naturaleza de las *relaciones entre las variables*. Todas las investigaciones cuantitativas, es decir, las que no son simplemente descriptivas, se interesan por las relaciones. Una relación o asociación, se da cuando una variable varía sistemáticamente ante otra variable. Esto puede lograrse tanto por la comparación de grupos diferentes como por correlaciones.

La figura 7.3 ilustra una relación establecida por la comparación de diferentes grupos. Las variables de interés son el curso y la autoestima. Existe una relación entre el curso y la autoestima porque disminuye progresivamente el número de estudiantes con una autoestima alta a medida que se aumenta de curso. En este caso, la relación es negativa, ya que en tanto aumenta el curso, el número de estudiantes con una elevada autoestima desciende. Podría examinarse esta misma pregunta de investigación con el cálculo de un coeficiente de correlación entre curso y puntuación de autoestima.



Fuente: McMillan, J. H. (2000), *Educational Research: Fundamentals for the Consumer*. 3.^a ed., New York: Longman, p. 180.

Las relaciones son importantes para nuestra comprensión de la enseñanza y del aprendizaje por varias razones. Primero, las relaciones nos permiten hacer una primera identificación de las posibles causas de resultados educativos importantes. Segundo, las relaciones nos ayudan a identificar variables que necesitaremos para posteriores investigaciones. Tercero, las relaciones nos permiten pronosticar una variable a partir de otra. Como veremos más tarde en este capítulo, la predicción es necesaria para una gran variedad de propósitos.

INVESTIGACIÓN COMPARATIVA

La finalidad de los estudios comparativos es investigar la relación de una variable con otra, examinando simplemente si el valor de la variable dependiente en un grupo es diferente del valor de la variable dependiente en el otro grupo. En otras palabras, la **investigación comparativa** examina las diferencias entre dos o más grupos de una variable.

Ejemplos de preguntas acerca de diferencias son los siguientes:

¿Existen diferencias entre los alumnos de octavo curso y los de noveno curso en sus actitudes hacia la escuela?

¿Cuál es la diferencia entre los estudiantes que asisten a los colegios privados y los que acuden a los colegios públicos?

¿Cuál es la diferencia entre los profesores principiantes y los experimentados?

¿Hay diferencia entre las puntuaciones de autoestima de los estudiantes de segundo, tercero y cuarto curso?

En cada caso, el investigador hace una comparación basada en datos descriptivos.

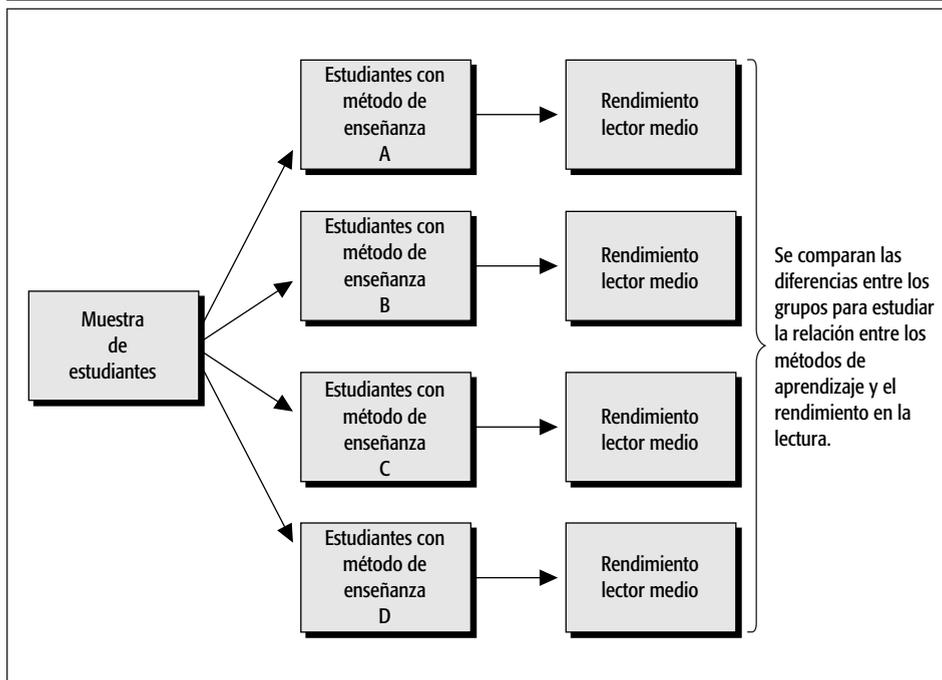
Un ejemplo sencillo de investigación comparativa es un estudio de la relación entre el género y las notas escolares. Puede compararse una muestra de las notas de las estudiantes femeninas con las de una muestra de estudiantes masculinos. Los resultados muestran cómo las diferencias en una variable, el género, se relacionan con las diferencias en otra variable, las notas. Si los resultados muestran que las mujeres tienen una media más alta de calificaciones (¡lo cual no es sorprendente!), esto indica que existe una relación entre las dos variables. Hay que señalar, sin embargo, que esto no es una *relación* causal. Aunque podemos predecir las notas conociendo el género, no podemos saber por qué el ser hombre o mujer afecta a las notas.

Otro ejemplo es un estudio de la relación entre los métodos de aprendizaje y el rendimiento. Suponga que tenemos cuatro tipos o categorías de métodos de aprendizaje y una medida de comprensión lectora. Puede obtenerse una muestra de estudiantes representativos de cada método de aprendizaje y compararse la

puntuación media en lectura para cada grupo. Si descubrimos que algunos métodos están asociados a un rendimiento mayor en la lectura, entonces hay una relación, aunque, como en el ejemplo del género y las notas, no puede concluirse que los métodos sean la causa del rendimiento. Este estudio hipotético se representa gráficamente en la figura 7.4.

Un buen ejemplo de una investigación comparativa publicada que investigó diferencias es un estudio del tiempo asignado a la enseñanza para los estudiantes de educación especial de primaria (Ysseldyke, Thurlow, Christenson y Weiss, 1987). Este estudio se llevó a cabo debido a que existían escasos conocimientos de investigaciones previas acerca de cómo se debía asignar el tiempo de aprendizaje a los estudiantes discapacitados. También existía escasa evidencia sobre si la asignación temporal debía variar según las diferentes categorías de discapacidad. De este modo, este estudio fue comparativo en el sentido de que investigó las diferencias entre las condiciones de discapacidad en la asignación de tiempo. La pregunta de investigación de este estudio que aparece en el extracto 7.1 ilustra la investigación comparativa.

FIGURA 7.4: DIAGRAMA DE UN ESTUDIO DE RELACIÓN POR COMPARACIÓN DE GRUPOS DIFERENTES



Fuente: McMillan, J. H. (2000), *Educational Research: Fundamentals for the Consumer*. 3.^a ed., New York: Longman, p. 181.

EXTRACTO 7.1**INVESTIGACIÓN DESCRIPTIVA COMPARANDO DIFERENCIAS**

¿Hasta qué grado existen diferencias entre la cantidad de tiempo asignada a la enseñanza de determinadas asignaturas específicas para los alumnos con dificultades de aprendizaje, con problemas emocionales/conductuales, con retraso mental educable y sin discapacidad?

Fuente: Ysseldyke, J. E, Thurlow, M. L., Christenson, S. L., y Weiss, J. (1987), «Time allocated to instruction of mentally retarded, learning disabled, emotionally disturbed, and nonhandicapped students», *The Journal of Special Education*, 21, 43–55.

El extracto 7.2 ilustra un ejemplo de investigación comparativa que investiga relaciones. En este estudio los investigadores entrevistaron a profesores de colegios públicos y privados católicos para investigar las diferencias en la orientación de la organización de valores, en el compromiso del profesor y en la satisfacción laboral. De esta forma, la pregunta es: ¿se observa una relación (diferencia) entre el tipo de colegio (público o privado) y la transmisión de valores, el compromiso del profesor y la satisfacción laboral?

El extracto 7.3 muestra cómo se emplea el término *comparativo* para describir la investigación.

El extracto 7.4 es un ejemplo del modo en que se presentan los datos descriptivos y comparativos en las revistas especializadas. Este estudio resumía las respuestas de los niños a la pregunta de por qué tuvieron éxito o fracasaron en el colegio. Esencialmente era una descripción de las respuestas de los niños. Puesto que para generar las respuestas se utilizan tanto las causas de éxito como de fracaso, la investigación también examina las diferencias de las causas del éxito comparadas con las de fracaso.

EXTRACTO 7.2**INVESTIGACIÓN DESCRIPTIVA INVESTIGANDO RELACIONES**

En este estudio investigamos la relación entre la orientación de valores y dos variables, el compromiso profesional y la satisfacción laboral, entre profesores de colegios públicos y privados... los resultados de este estudio sugirieron que (a) los colegios privados exhiben un orientación más basada en reglas; y (b) los colegios con una orientación de valores más reglada tenían un compromiso profesional por parte del profesor significativamente más alto que los colegios con una orientación de valores más utilitaria.

Fuente: Reyes, P. y Pounder, D. G. (1993), «Organizational orientation in public and private elementary schools.», *Journal of Educational Research*, 87, 86–93.

EXTRACTO 7.3 INVESTIGACIÓN COMPARATIVA

El propósito era identificar variables que pueden promover ambientes de aprendizaje de ciencias eficaces para todos los estudiantes. Específicamente, las comparaciones se hicieron entre las percepciones de estudiantes masculinos y femeninos y estudiantes blancos y negros dentro de las mismas clases. Además, se compararon las percepciones de los ambientes de aprendizaje por parte de los estudiantes con profesores hombres o mujeres, así como con profesores blancos y negros.

Fuente: Huffman, D., Lawrenz, F. y Minger, M. (1997), «Within-class analysis of ninth-grade science students' perceptions of the learning environment», *Journal of Research in Science Teaching*, 34, 791–804.

EXTRACTO 7.4 RESULTADOS DESCRIPTIVOS Y COMPARATIVOS

Los datos de la tabla 1 indican que los niños fueron capaces de identificar las variables que afectaban a su éxito o a su fracaso. Sus atribuciones iniciales estuvieron principalmente referidas a la tarea (del 46% al 58% dijeron que las palabras eran fáciles). La siguiente causa más común para conseguir éxito era su propio esfuerzo (40% de las respuestas). Cuando se les pedía una segunda respuesta, los sujetos, de manera uniforme, dividían sus respuestas entre los cuatro tipos de atribuciones. A partir de las respuestas totales, cuando tenían éxito en la lectura de una palabra, era más probable que lo atribuyeran a su esfuerzo (33%) o a la dificultad de la tarea (37%). Cuando fallaban en la lectura de una palabra era más probable que atribuyeran su fracaso a las características de la tarea (40%).

TABLA 1: Porcentajes de niños que nombraron varios atributos como respuesta a por qué tuvieron éxito (E) o fracasaron (F)

Tipos de atribuciones	Condición de la pregunta					
	Inicial ^a		Segunda ^b		Total	
	E	F	E	F	E	F
Capacidad	.10	.02	.22	.05	.16	.03
Esfuerzo	.40	.12	.25	.15	.33	.14
Dificultad tarea	.46	.58	.28	.22	.37	.40
Suerte	.02	.15	.15	.20	.08	.18
Otras	.02	.12	.10	.38	.06	.25

^a Inicial incluye a los niños que contestaron espontáneamente y a los que se les dio elección de respuesta si no respondían de forma espontánea.

^b Segunda es una segunda atribución que dio el niño.

^c Otros se compone de «no sé», «supongo» y la falta de respuesta.

Nota: N = 40.

Fuente: Cauley, K. M., y Murray, F. B. (1982), «Structure of Children's Reasoning About Attributes of School Success and Failure», *American Educational Research Journal*, 19, 473–480. Copyright © 1982, American Educational Research Association, Washington, DC. Reproducido con permiso.

Al igual que los estudios descriptivos, la investigación comparativa necesita proporcionar una descripción clara de los sujetos, la instrumentación y los procedimientos. Como sucede en la investigación descriptiva, tampoco pueden hacerse conclusiones causales. Lo mejor que puede sacarse como conclusión es que existe una diferencia o una relación. Este principio se pasa por alto fácilmente, ya que algunos estudios comparativos parecen establecer una conexión causal entre las variables dependiente e independiente. Por ejemplo, suponga que se describe que los alumnos de colegios privados superan a los de los colegios públicos. Es tentador concluir que la razón, o la causa, de la diferencia es la naturaleza del colegio. Sin embargo, hay otras muchas explicaciones posibles, como por ejemplo, las diferencias en la implicación de los padres o la situación socioeconómica de los alumnos.

A menudo, los resultados de la investigación comparativa se presentan con gráficos. Como apuntamos en el capítulo 5, algunas veces, una observación no coincide totalmente con los datos reales. Cuando se interpretan esos gráficos, hay que asegurarse de examinar las cifras y escalas reales en las que se basa el gráfico y estar alerta a las distorsiones que se pueden producir.

INVESTIGACIÓN CORRELACIONAL

El capítulo 5 describía el coeficiente de correlación. En este capítulo manejamos la idea básica de correlación como una forma de conceptualizar los problemas de investigación. Primero, presentamos estudios de relación simple; luego, de investigación predictiva multifactorial más compleja, seguida por la cautela a la hora de interpretar correlaciones.

Estudios correlacionales bivariados

En un estudio correlacional bivariado, los investigadores obtienen puntuaciones a partir de dos variables para cada sujeto y entonces emplean los pares de puntuaciones para calcular un coeficiente de correlación. El término **bivariado** u *orden cero* significa que dos variables están correlacionadas. Las variables se seleccionan porque la teoría, la investigación o la experiencia sugieren que pueden estarlo. Se selecciona, luego, una muestra y se recogen los datos. En los estudios correlacionales, es importante que los investigadores seleccionen a los sujetos para proporcionar un rango de respuestas sobre las variables. Si los sujetos son homogéneos con respecto a cada variable, es poco probable una relación entre éstas. De forma similar, es importante seleccionar instrumentos/ pruebas que sean fiables y que proporcionarán un rango de respuestas. Pueden emplearse varios métodos de instrumentación, incluyendo tests, cuestionarios, entrevistas

u observaciones. Independientemente del tipo de pruebas, lo mejor es realizar un estudio piloto o manejar datos previos de sujetos similares para asegurar la fiabilidad y la variabilidad en las respuestas. Por ejemplo, a menudo, es difícil relacionar las evaluaciones de los estudiantes sobre los profesores con otras variables debido a un efecto de techo en tales evaluaciones; lo que supone que la mayoría de los profesores obtienen evaluaciones elevadas. De forma similar, es poco probable que las puntuaciones normativas de rendimiento de los estudiantes con altas capacidades correlacionen con otras variables, porque las puntuaciones pueden tener un rango restringido (trataremos el rango restringido, con más detalle, más adelante).

En algunos estudios de relación pueden presentarse las correlaciones bivariadas de distintas variables. De hecho, una ventaja de la investigación correlacional es que permite el estudio simultáneo de varias variables. Sin embargo, para algunos investigadores, sin una justificación razonable, es posible medir un gran número de variables para encontrar algunas relaciones significativas. Esto se denomina aproximación *escopeta* y se emplea de forma inapropiada con la esperanza de que alguna de las muchas correlaciones calculadas indique relaciones significativas.

Un ejemplo de cómo se presentan los resultados en un estudio correlacional bivariado se ilustra en el extracto 7.5. En este estudio las correlaciones bivariadas se calcularon para explorar las relaciones entre las actitudes de los adolescentes y sus madres hacia la ciencia, el valor intrínseco de la ciencia, el apoyo de los compañeros, las actividades disponibles de ciencias, de curso, de nota media de ciencias del curso y la preferencia sobre futuras carreras de ciencias.

Estudios predictivos

En la educación existen muchas situaciones en las que necesitamos realizar predicciones. Los profesores vaticinan las reacciones de los estudiantes al hacer ciertas tareas. Los directores prevén el comportamiento del profesor sobre la base de los criterios usados para la evaluación de la eficacia de los profesores. Los profesores aconsejan a los estudiantes que se centren en especialidades particulares de acuerdo con su interés laboral o los tests de aptitud. Los estudiantes se seleccionan para los programas especiales debido a que los profesores predicen que podrían hacerlo mejor que otros estudiantes.

Realizamos **estudios predictivos** para proporcionar una estimación más exacta del pronóstico. Suponga que es el director de admisiones de una universidad pequeña y selectiva. Cada año, un gran número de estudiantes solicitan ingresar en ella, muchos más de los que pueden ser admitidos. ¿Cómo decidiría qué estudiantes deberían entrar? Puede extraer nombres al azar de un sombrero, pero entonces serían admitidos algunos alumnos que no acabarían sus estudios, mientras

EXTRACTO 7.5**PRESENTACIÓN DE LOS RESULTADOS DE UNA CORRELACIÓN BIVARIADA****Correlaciones de orden cero entre las variables del contexto motivacional percibido y los índices de alineación**

Variables independientes	1	2	3	4	5	6	7	8	9
1. Ciencias GPA	1.0								
2. Curso en el colegio	.02	1.0							
3. Apoyo de los amigos	.16	.05	1.0						
4. Número de actividades de ciencias/matemáticas	.21*	-.15	.20*	1.0					
5. Número de actividades que no son de ciencias	.21*	.02	.08	.23**	1.0				
6. Percepción de las madres sobre las habilidades de ciencias de los niños	.56**	-.05	.10	.17*	.07	1.0			
7. Valoración de las madres de las ciencias para las mujeres	.13	-.01	.09	.19*	.03	.30**	1.0		
8. Interés de los adolescentes en la biología	.28**	.02	.19*	.36**	.14	.29**	.16	1.0	
9. Interés de los adolescentes en la física	.29**	.08	.08	.29**	.23**	.39**	-.24**	.47**	.10

* $p < .05$. ** $p < .01$.

Fuente: Jacobs, J. E., Finken, L. L., Griffin, N. L., y Wright, J. D. (1998), «The career plans of science-talented rural adolescent girls», *American Educational Research Journal*, 35 (4), 681–704.

que otros con buenas aptitudes serían rechazados. Usted comprende que lo haría mucho mejor si pudiese *predecir*, basándose en características ya establecidas, qué estudiantes tienen más probabilidad de tener éxito. Dado que parece razonable que el rendimiento previo predecirá el rendimiento posterior, puede considerar si existe una correlación entre la nota media del instituto (GPA) (previo) y la GPA de la universidad (rendimiento posterior). Cuando descubre que estas dos variables se correlacionan con un .70, entonces, posee información de utilidad para seleccionar a los estudiantes. O, lo que es lo mismo, es más probable que los alumnos con una GPA alta en el instituto tengan una GPA alta en la universidad que los estudiantes con GPA bajas en el instituto.

En este caso, la GPA del instituto es una **variable predictora** y la GPA de la universidad es una **variable criterio**. La variable predictora se determina *antes* que la variable criterio. De hecho, en los estudios de predicción, los resultados como la nota media, estudiantes con fracaso académico, éxito como líder o gestor, efectividad de una droga para curar una enfermedad etc., se relacionan con los comportamientos que ocurrieron previamente al criterio. Para hacer esto, es

necesario tener datos de los sujetos que abarquen un cierto periodo de tiempo. Esto puede hacerse de forma retrospectiva a través de registros de los sujetos, o bien, puede realizarse longitudinalmente mediante la recogida, en primer lugar, de datos de la variable predictora, esperar un periodo de tiempo apropiado y, después, recoger los datos de la variable criterio. Por ejemplo, suponga que necesita seleccionar los mejores profesores para un nuevo departamento de su escuela. Esencialmente, está prediciendo que los nuevos profesores que elige serán eficaces. En su región todos los probables profesores hacen los Exámenes de Profesor (EP)¹, por lo tanto, es capaz de estudiar la relación predictiva del EP con la eficacia del profesor (medida por medio de las evaluaciones del director y del inspector). Una vez que ha establecido el poder predictivo del EP con un grupo de profesores, debería «comprobar» la relación predictiva con otro grupo de nuevos profesores. En la «comprobación» de la predicción, el investigador utiliza los valores de las variables predictoras a partir de un nuevo grupo de futuros profesores; en este ejemplo, las puntuaciones actuales del EP, valorando entonces cada puntuación por un factor calculado a partir de la ecuación predictiva original. Esto indicará si, como bien puede esperar el investigador, el EP predice la efectividad del profesor. La relación «comprobada» será menor que la originalmente empleada para establecer la ecuación predictiva.

El extracto 7.6 proporciona un buen ejemplo de una investigación predictiva. En este estudio se utilizan el éxito de los estudiantes en programas para superdotados como variable criterio y las puntuaciones WISC-R, las recomendaciones de los profesores, las notas y las puntuaciones de los tests de rendimiento como variables predictoras. Observe que un grupo de estudiantes se utiliza para establecer la relación predictiva y otro grupo para «comprobarla».

Puede ocurrir que, si maneja varias variables predictoras, pueda realizar una predicción más exacta. Suponga que en nuestro ejemplo de la predicción de la enseñanza eficaz, también tenemos información de la GPA de la universidad, referencias, los resultados de una entrevista y un enunciado por cada candidato además de las puntuaciones del EP. Cada sujeto recibiría una puntuación para cada variable (las referencias, los resultados de la entrevista y los enunciados se juzgarían y puntuarían de acuerdo con un sistema de valoración). La totalidad de las variables predictoras puede combinarse para formar lo que se denomina una **ecuación de predicción de regresión múltiple**. Esta ecuación añade de forma conjunta el poder predictivo de varias variables independientes. Cada variable predictora podría representarse por X_1 , X_2 , X_3 , etc. y la variable criterio por Y . De esta forma, en nuestro ejemplo:

$$Y = X_1 + X_2 + X_3 + X_4 + X_5$$

¹ Equivalente en España a las oposiciones de Magisterio.

Por tanto

Y = efectividad en la enseñanza

x_1 = puntuación EP

x_2 = GPA de la universidad

x_3 = evaluación de las referencias

x_4 = evaluación de la entrevista

x_5 = evaluación de los enunciados del candidato

EXTRACTO 7.6 **INVESTIGACIÓN PREDICTIVA**

Este estudio fue realizado en dos fases. La fase inicial del estudio implicó a 120 alumnos de primaria de entre 6 y 11 años. Estos estudiantes tenían diferentes grados de éxito en programas para superdotados... La segunda fase de la investigación implicó una selección aleatoria de 41 sujetos adicionales del mismo programa... En la segunda fase, se intentó, a partir de los resultados de los datos de la primera fase, predecir los estudiantes que tendrían éxito o fracasarían en el programa.

Fuente: Lustberg, R. S., Motta, R., y Naccari, N. (1990), «A model using the WISC-R to predict success in programs for gifted students», *Psychology in the Schools*, 27, 126–131.

Para obtener una puntuación de eficacia del profesor, deben colocarse los valores de cada una de las cinco variables predictoras en la ecuación y cada uno debe ponderarse por un número, llamado **coeficiente de regresión**, para determinar la contribución de cada factor a la predicción de la eficacia del profesor. Ya que las unidades de cada variable predictora son diferentes (las valoraciones podrían abarcar de 0 a 10, las GPA de 0 a 4) los coeficientes de regresión no pueden compararse directamente en la ecuación. Para comparar el poder predictivo de las variables, los coeficientes de regresión se convierten en **pesos beta**, que sí pueden compararse directamente. De esta forma, es posible comparar la contribución relativa de cada variable en nuestro ejemplo. Si el peso beta para EP es .32 y el peso beta para GPA es .48, entonces la GPA contribuye más que el EP en la predicción de la eficacia del profesor. El efecto combinado de las variables independientes, en términos de poder predictivo, sobre la variable dependiente, se representa por R, el coeficiente de correlación múltiple. El **coeficiente de correlación múltiple** puede interpretarse como una correlación simple de todas las variables independientes juntas con la variable dependiente.

Cuando se planifica un estudio predictivo, el investigador debe tener en cuenta varios factores que afectarán a la exactitud de la predicción. Uno es la fiabilidad de la puntuación de las variables predictoras y de criterio. Unas puntuaciones más fiables darán como resultado predicciones más exactas. Otro factor es el periodo

de tiempo entre las variables predictoras y de criterio. En la mayoría de los casos, las predicciones que implican espacios cortos de tiempo (por ejemplo, semanas o meses) son más exactas que aquellas en las que hay largos espacios de tiempo entre las variables predictoras y de criterio (por ejemplo, años). Esto es debido al principio general de que la correlación entre dos variables disminuye conforme aumenta el espacio de tiempo entre ellas; además, cuanto más tiempo pase, hay más probabilidad de que otras variables afecten a la variable criterio, lo que podría disminuir la predicción. Finalmente, algunas variables criterio, como por ejemplo, éxito en el colegio, liderazgo y enseñanza eficaz, son más difíciles de predecir porque están influidas por muchos factores. Las variables criterio relativamente simples, como por ejemplo, el éxito en el próximo examen de matemáticas o ser admitido en, al menos, una universidad, son mucho menos difíciles de predecir.

Se debe tener en cuenta que una regresión múltiple es un procedimiento de análisis de datos versátil que se emplea para muchos tipos de estudios diferentes, no sólo en la investigación predictiva. Puede usarse en estudios comparativos e, incluso, experimentales y con variables de naturaleza nominal, ordinal o de intervalo. Normalmente la regresión múltiple se emplea siempre que los investigadores están interesados en estudiar la relación de varias variables independientes combinadas con una variable dependiente. La regresión múltiple también permite a los investigadores «controlar» variables seleccionadas para determinar la relación entre las variables independientes restantes y la variable dependiente. Por ejemplo, si un estudio investiga la relación entre el tamaño de la clase y el rendimiento, primero puede «introducirse» en la regresión una medida de la situación socioeconómica para que pueda «ser controlada» y permitir una conclusión sobre la relación sin tener que preocuparse de que la situación socioeconómica se confunda con el tamaño de la clase (por ejemplo, las clases menos numerosas tienen alumnos con unas circunstancias socioeconómicas más altas). Es como si se pudiera eliminar la variabilidad debida a la situación socioeconómica, lo que permite relacionar la variabilidad restante con el tamaño de la clase. El extracto 7.7 muestra cómo las variables del entorno del estudiante se «controlan» en un estudio de la relación entre las actividades extraescolares y el rendimiento académico.

El extracto 7.8 proporciona un ejemplo del empleo de la correlación múltiple. En este estudio se predijo el síndrome de *burnout* del profesor por medio de once variables independientes. La tabla resume la relación predictiva de las once variables con los tres índices de *burnout*: desgaste emocional, despersonalización y satisfacción personal. De las once variables, sólo se han incluido en la tabla aquellas que se introdujeron primero (sexo y edad) o las que afectan de forma significativa a la relación global. En las regresiones múltiples, pueden introducirse, en primer lugar, algunas variables para el control de ciertos factores, es decir, para eliminar esos factores como influencias a la relación. En este caso, los investigadores querían controlar el efecto de la edad y el sexo para advertir cuánto podrían predecir las otras variables.

EXTRACTO 7.7**ANÁLISIS DE REGRESIÓN MÚLTIPLE PARA «CONTROLAR» VARIABLES SELECCIONADAS**

Cada regresión utilizaba una de las tres medidas de rendimiento como variable criterio. En los tres análisis, se introdujeron, simultáneamente, las variables de contexto de cinco estudiantes como el primer grupo de variables de predicción de las medidas de rendimiento. Las variables de contexto fueron el género, el curso, la raza, la elección de comedor gratuito y si un adulto estaba con ellos o no después de clase. Como segundo paso del análisis jerárquico de regresión múltiple, se añadieron las cinco variables de las actividades extraescolares como un grupo para determinar si explicaban una cantidad significativa de variabilidad en las medidas de rendimiento por encima de las variables contextuales por sí solas.

Fuente: Cooper, H., Valentine, J. C. y Nye, B. (1999), «Relationships between five after-school activities and academic achievement», *Journal of Educational Psychology*, 91 (2), 369-378.

Aunque la inmensa mayoría de la investigación correlacional se ocupa sólo de relaciones, algunas técnicas estadísticas recurren a correlaciones múltiples para investigar las cuestiones de causa y efecto. Una técnica, **el análisis de sendas**, utiliza las correlaciones de varias variables para estudiar patrones causales. Un «modelo» causal se establece, basado en la teoría, para mostrar a través de flechas los patrones causales. Las correlaciones entre las variables del modelo proporcionan una evidencia empírica de los enlaces causales propuestos. Una técnica relativamente nueva, el *modelo de ecuaciones estructurales* o *modelo de rasgo* o *variable latente*, es más potente que el *path analysis* ya que las puntuaciones tienden a ser más fiables y las inferencias más válidas. LISREL es un tipo común de formulación de ecuación estructural. Aunque estas técnicas son útiles para examinar relaciones causales, son estadísticamente sofisticadas y difíciles de utilizar. Además, poseen las mismas limitaciones fundamentales que todos los datos correlacionales. Las variables no medidas relacionadas tanto con las variables dependientes como con las independientes son siempre una fuente de explicaciones causales potenciales y la dirección no siempre está clara.

Interpretación de la investigación correlacional

Los coeficientes de correlación se emplean generalmente en la investigación y parecen ser índices de relación simples y sencillos. Sin embargo, es necesario conocer varias limitaciones en el empleo de la correlación. La mayoría hacen referencia a la sobrevaloración. En esta sección presentamos algunos principios importantes que nos ayudarán a entender el significado de los resultados de una investigación correlacional.

EXTRACTO 7.8**INVESTIGACIÓN DE CORRELACIÓN MÚLTIPLE**

Los resultados presentados en la tabla 2 apoyan las impresiones previas de que variables que contribuyen al *burnout* tienen efectos diferentes sobre los aspectos de agotamiento emocional, despersonalización y realización personal (Maslach y Jackson, 1981; Schwab y Iwanicki, 1982a). Por ejemplo, después de controlar la edad y el sexo, el conflicto de desempeño de función explicaba el mayor porcentaje de varianza en el agotamiento emocional y la despersonalización (24% y 12% respectivamente); mientras que la autonomía afectaba a la realización personal (12%). Cinco de los predictores supuestos tenían influencia para una varianza significativa en el agotamiento emocional, cuatro para la despersonalización y sólo dos para la realización personal. El único predictor relacionado significativamente con todos los aspectos fue el apoyo social de los colegas. Como indicaron los pesos beta, las relaciones entre las condiciones de organización significativas y los aspectos del *burnout* estaban en la dirección supuesta, y ya habían sido presentadas en investigaciones previas (Maslach y Jackson, 1981; Schwab y Iwanicki, 1982a).

TABLA 2: Análisis de regresión múltiple por pasos para la relación entre las variables significativas personal/organizativa y tres aspectos del *burnout* cuando se controlan la edad y el sexo (N = 339)

Paso variable introducida	Incremento			Para R ²	
	R	R ²	R ²	Beta	F**
Agotamiento emocional					
1 Sexo y edad ^a	.120	.014	.014		NS ^b
2 Conflicto de funciones	.493	.243	.229	.48	35.17
3 Expectativas	.537	.288	.059	.26	33.14
4 Apoyo social de los colegas	.553	.306	.018	2.16	28.79
5 Castigo eventual	.562	.315	.009	.10	24.95
6 Ambigüedad de funciones	.570	.325	.010	2.13	22.27
Despersonalización					
1 Sexo y edad ^a	.285	.081	.081		14.56
2 Conflicto de funciones	.352	.124	.043	.21	15.50
3 Participación en la toma de decisiones	.374	.140	.016	2.14	13.26
4 Castigo eventual	.393	.155	.015	.12	11.92
5 Apoyo social de los colegas	.407	.166	.011	2.12	10.76
Realización personal					
1 Sexo y edad ^a	.192	.037	.037		6.33
2 Autonomía	.349	.122	.045	.29	15.14
3 Apoyo social de los colegas	.366	.134	.012	.12	12.62

^a Las variables antecedentes de sexo y edad se introdujeron dentro del análisis por pasos en el paso 1. El resto de las variables se introdujeron por pasos.

^bNS = no significativo

**p < 0.01.

Fuente: Schwab, R. L., Jackson, S. E. y Schuler, R. S. (1986), «Educator burnout: Sources and consequences», *Educational Research Quarterly*, 10, (3). Copyright © 1986 University of Southern California. Reproducido con permiso.

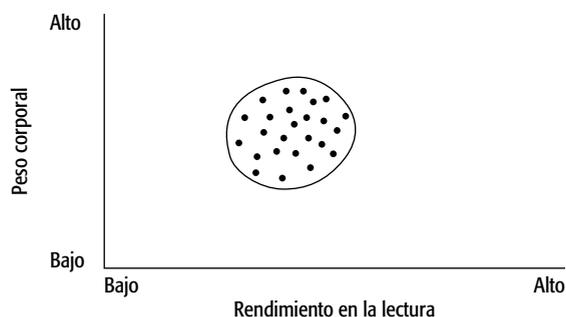
Correlación y causalidad. Puede estar familiarizado con el famoso mandamiento «nunca infiera causalidad a partir de una correlación». Éste es un principio que supuestamente se graba a fuego en las mentes de los estudiantes y por un buen motivo, ya que probablemente es el principio de medidas de relación más transgredido.

Existen dos razones para que no se infiera causalidad de la correlación; primero, una relación entre X e Y puede ser alta, pero no hay forma de saber si X causa Y o Y causa X y, segundo, puede haber variables no medidas que afecten a la relación. Con respecto a la dirección de la posible causalidad, considere este ejemplo: un investigador encuentra una correlación positiva alta entre la autoestima y el rendimiento. ¿Significa esto que un incremento en la autoestima causará un aumento en el rendimiento? Tal vez, pero sería igualmente probable que un incremento del rendimiento tuviera como resultado una autoestima más elevada. Si encontramos que los índices de fracaso escolar están asociados de forma negativa con los sueldos de los profesores, ¿deberíamos asumir que aumentar el sueldo de los profesores tendrá como consecuencia un menor fracaso escolar?

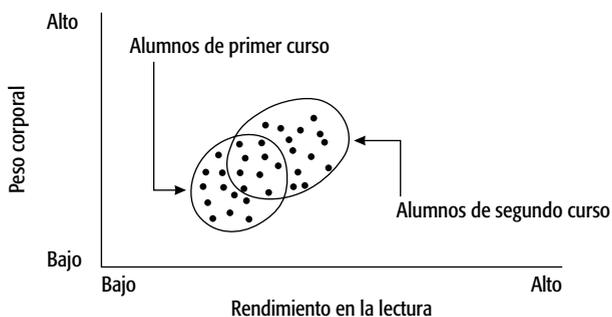
Las variables no tenidas en cuenta también son importantes para la interpretación de las correlaciones. Digamos que se observa una relación positiva entre los colegios religiosos y la honestidad. Aunque los colegios pueden ayudar a fomentar una mayor honradez, hay muchas otras variables, como por ejemplo, las creencias familiares, las actitudes paternas y los métodos de disciplina, que estarían más relacionados. O, por ejemplo, ¿qué hay acerca del hallazgo de que escuelas que gastan más por alumno alcanzan un mayor rendimiento? Sería un error inyectar dinero en las escuelas pobres con la esperanza de que esto provoque un mejor rendimiento, ya que es más probable que el entorno familiar, una variable no medida, sea un factor más importante en el rendimiento que el gasto por alumno.

Estas dos limitaciones parecen bastante sencillas, pero aun así las correlaciones todavía son malentendidas. Considere el *hecho* de que hay una fuerte relación positiva entre el peso corporal y el rendimiento en la lectura. ¿Increíble? Examine la explicación que sigue (adaptada de Helperin, 1978).

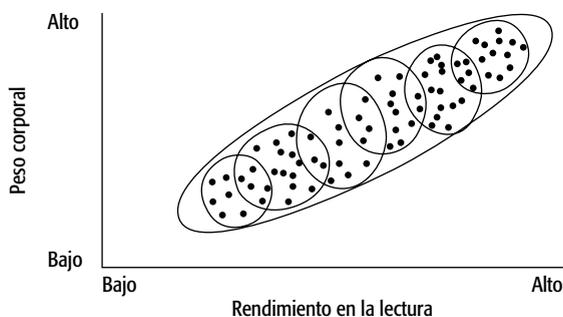
1. Representar gráficamente el peso corporal y las puntuaciones de un grupo de alumnos de primer curso.



2. A continuación, añade las puntuaciones de los alumnos de segundo curso:



3. Finalmente, añade las puntuaciones de los alumnos de tercer a sexto curso:



¡Evidente! Hemos obtenido una relación positiva entre el peso corporal y el rendimiento en la lectura mediante la combinación de una serie de correlaciones cercanas a cero. ¿Por qué? Porque una tercera variable que no se incluyó, la edad, estaba relacionada con el peso corporal y, obviamente, existe una relación positiva entre la edad y el rendimiento en la lectura. ¡Si el lector cree que esta correlación implicaba causalidad, entonces podría mejorarse el rendimiento en lectura engordando a los alumnos! (O, la mejora del rendimiento en lectura podría dar como resultado estudiantes más gordos).

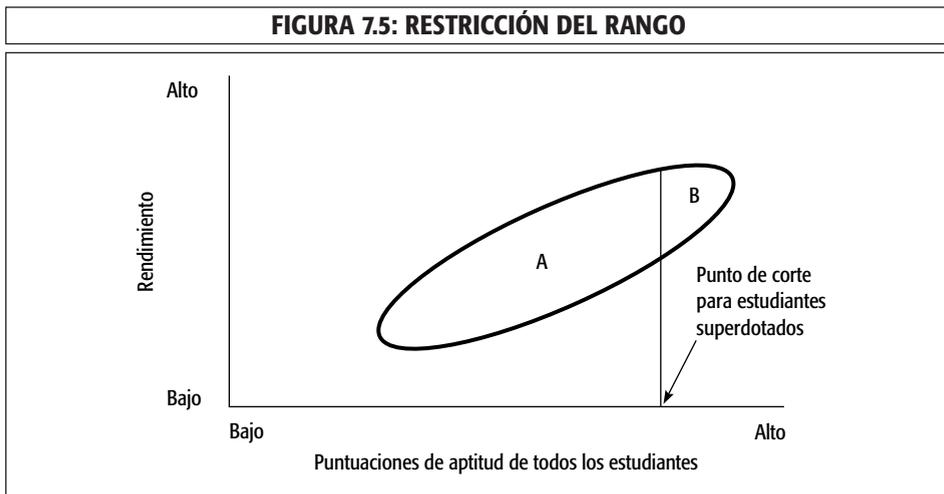
Correlaciones espúreas. Cuando una correlación representa de forma excesiva o exigua la relación real se denomina una **correlación espúrea**. Las relaciones espúreas que sobreestiman las relaciones se obtienen si hay una variable común que forma parte, tanto de las variables dependientes como de las independientes. Por ejemplo, si un investigador posee datos de un pretest y de un postest y mide la relación entre las puntuaciones del postest y el aumento de puntuación entre el pretest y el postest, la correlación será espúreamente alta, porque la puntuación del postest está incluida en ambas variables. Obviamente, cuando algo se corre-

laciona consigo mismo la relación será muy alta. De forma semejante, si hay una tercera variable no medida, es decir, común a ambas variables como en nuestro ejemplo sobre el rendimiento lector y el peso corporal, la correlación será espúreamente alta. Un resultado así se obtendría cuando la altura se correlacionara con el peso ya que el tercer factor, la edad, es común.

Las correlaciones obtenidas a partir de dos puntuaciones que no son perfectamente fiables tendrán como resultado coeficientes más bajos que la verdadera relación entre las puntuaciones. Esta disminución de los coeficientes se conoce como **atenuación** y, de forma ocasional, los investigadores calcularán una corrección para la atenuación para estimar lo que podría haber sido la correlación si las puntuaciones hubiesen sido más fiables. Un investigador empleará con más frecuencia esta correlación en estudios piloto, en los que las puntuaciones que se manejan poseen una baja fiabilidad.

Otra situación en la que el coeficiente de correlación es menor que la relación real es un caso en el que el rango de puntuaciones sobre una de las variables está limitado a una representación de sólo una parte de la distribución total de estas variables. Este problema se llama **restricción del rango** y tiene como resultado una disminución en la correlación. Suponga, por ejemplo, que un investigador quiere estudiar la relación entre las puntuaciones de un test de aptitud y el rendimiento de los estudiantes en un programa para superdotados. La figura 7.5 muestra por qué el investigador encontrará, probablemente, una pequeña relación positiva entre estas dos variables. Hay que observar que la figura ilustra la aptitud hipotética y las puntuaciones de rendimiento para todos los estudiantes (A y B), junto con la muestra de estudiantes del programa para superdotados (B).

Por lo tanto, en este caso, el rango de puntuaciones de aptitud esta limitado, o «restringido», a una pequeña parte del rango total de puntuaciones de aptitud.



Si se utiliza el rango total de puntuaciones, entonces la correlación es altamente positiva. La restricción del rango explica, en parte, la relación, normalmente modesta, entre los tests de admisión de las universidades y el rendimiento en la universidad: el rango está restringido a los estudiantes que tienen puntuaciones altas en los tests de admisión. Bajo ciertas condiciones, se puede aplicar una corrección para la restricción del rango para indicar cuál sería la correlación si hubiese sido mayor.

Otra situación que puede llevar a correlaciones espúreas es aquella en que los procedimientos de muestreo tienen como resultado una muestra más homogénea o más heterogénea de la que existe realmente en la población. Este sesgo en el muestreo debería llevar a una correlación espúrea alta, si la muestra es más heterogénea; o a una correlación espúrea baja para una muestra más homogénea. Por ejemplo, si un investigador, al estudiar la relación entre el comportamiento de enseñanza eficaz y el rendimiento de los alumnos, representa de forma excesiva a los estudiantes con alto rendimiento, obtendrá como resultado un grupo más homogéneo que la población, por lo que la correlación será espúreamente baja. Sin embargo, si la población estuviera mayoritariamente compuesta por alumnos de altas capacidades pero los procedimientos infravaloraron este dato, entonces la correlación sería espúreamente alta.

Por último, pueden aparecer correlaciones espúreas si se incluyen en los cálculos las puntuaciones extremas (*outliers*). Presentamos el concepto de puntuación extrema en el capítulo 5, en relación con la necesidad de crear gráficos de dispersión de los datos relacionales. La figura 7.6 muestra dos gráficos de dispersión para ilustrar el efecto de que una puntuación extrema puede, tanto aumentar como disminuir el coeficiente de correlación, de manera que la correlación no represente bien los datos. Se han calculado ambas correlaciones con y sin puntuaciones extremas. Como puede observar, la diferencia en los resultados obtenidos es drástica.

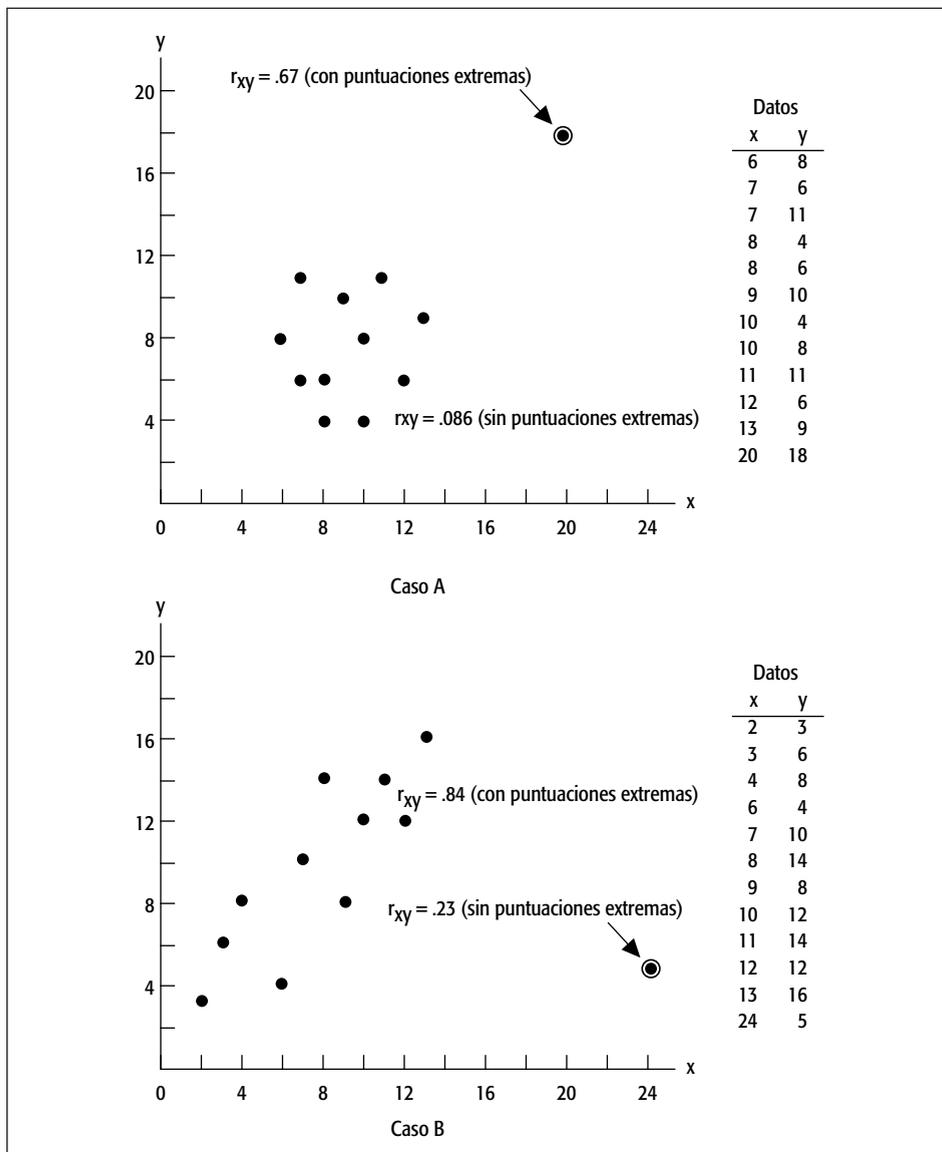
Tamaño de los coeficientes de correlación. Como ya se apuntó en el capítulo 5, las correlaciones como por ejemplo .86, .95 y $-.89$ son altas; .43, $-.35$ y .57 son moderadas; y .07, $-.01$ y .12 son bajas, pero estas cifras sólo indican la magnitud de la relación. Debido a que los coeficientes de correlación se expresan como decimales, es fácil confundir los decimales con porcentajes. El coeficiente es una forma matemática de expresar el grado en el que hay una covarianza entre las variables, no una indicación del grado en el que las variables comparten propiedades o características comunes. Para obtener una estimación de la proporción de la varianza que comparten o tienen en común las dos medidas, el coeficiente debe estar al cuadrado. Por ejemplo, una correlación de .40, elevada al cuadrado, indica que las variables tienen el 16% de la varianzas en común. En otras palabras, el 84% está sin explicar o sin predecir por la correlación de .40. Incluso para algunas correlaciones altas, como por ejemplo .70 y .80, el cuadrado de las

correlaciones tiene como resultado un grado moderado de varianza común (49% y 64%, respectivamente, de un total del 100%, lo que sería una relación perfecta). El índice que resulta de elevar al cuadrado la correlación se denomina **coeficiente de determinación**.

Otra consideración con respecto al tamaño de las correlaciones es que muchas correlaciones se denominan «significativas» incluso siendo bastante bajas (como por ejemplo .15, .08). Los investigadores emplean la palabra *significativo* en el contexto de las correlaciones para indicar que el coeficiente es estadísticamente diferente de cero (sin relación) con un nivel de confianza específico (el capítulo 9 trata en profundidad la significación estadística y el nivel de confianza). Si un estudio tiene un gran número de sujetos (más de 1.000), entonces, las correlaciones pequeñas pueden ser «significativas», pero sólo en un sentido estadístico. Sabemos que una simple correlación de .30, si es significativa, sólo explica el 9% de la varianza común, de modo que nuestra interpretación necesita reflejar el 91% de la varianza que no ha sido tomada en cuenta. Para la investigación en la que la predicción es la principal meta, las correlaciones bajas, aunque estadísticamente significativas, poseen una baja significación práctica. Generalmente, en los estudios que investigan sólo relaciones correlaciones bajas como .30 o .40 pueden ser de utilidad, pero en los estudios predictivos o en las estimaciones de fiabilidad y validez se necesitan correlaciones más altas. En la investigación educativa, si se presenta una correlación muy alta, como por ejemplo, 1.00, o .99, o .98, entonces el lector debería sospechar de la existencia de resultados espúreos, causados por una metodología, un diseño o un cálculo inadecuados. Como una ilustración final del poder de las correlaciones, observe los gráficos de dispersión de la figura 7.7. Estas figuras muestran la apariencia gráfica de las actuales correlaciones. Observe la puntuación 5.5 en el eje horizontal de la figura 7.7 y eche un vistazo a lo largo del rango de las puntuaciones que predice. Para el gráfico de dispersión (a), que tiene una correlación «alta» de .75, las puntuaciones van desde 3 hasta 8. Esto no es ni remotamente tan preciso como implicaría la correlación de .75.

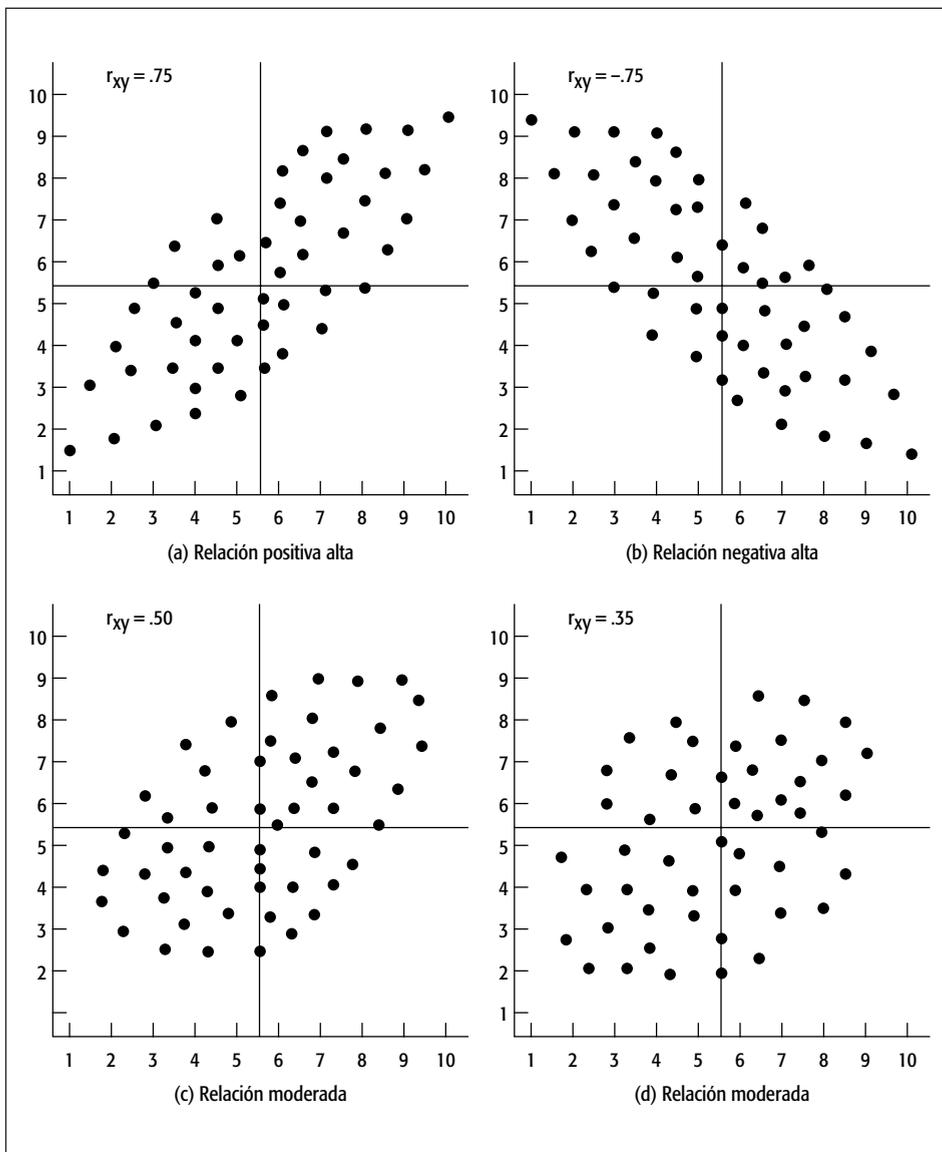
Finalmente, con respecto a la interpretación, la utilidad de las correlaciones varía dependiendo de si la investigación se centra en grupos o en individuos. Generalmente, es necesaria una correlación mayor para su empleo con individuos que con grupos. Para las correlaciones por debajo de .35, sólo se muestra una pequeña relación; esto es valioso en alguna investigación exploratoria, pero tiene un escaso valor en las predicciones relacionadas con individuos o grupos. En un rango medio, de .35 a .75, pueden hacerse predicciones grupales toscas; si pueden combinarse varias correlaciones moderadas, éstas pueden ser empleadas para realizar predicciones individuales. Por encima de .75 tanto las predicciones individuales como las de grupo son útiles con una medida de correlación simple.

FIGURA 7.6: EL EFECTO DE LAS PUNTUACIONES EXTREMAS EN LOS COEFICIENTES DE CORRELACIÓN



Fuente: Stevens, J, (1996), *Applied Multivariate Statistics for the Social Sciences*, 3.^a ed., Mahwah, NJ: Lawrence Erlbaum Associates, p. 16.

FIGURA 7.7: DIAGRAMAS DE DISPERSIÓN QUE INDICAN CORRELACIONES DE VARIOS TAMAÑOS (N = 50)



Fuente: Mehrens, W. A. y Lehmann, I. J. (1987), *Using Standardized Tests in Education*, 4.^a ed., White Plains, NY: Longman, p. 49.

INVESTIGACIÓN MEDIANTE ENCUESTA

En la investigación mediante encuesta, el investigador selecciona una muestra de entrevistados y les administra un cuestionario o realiza entrevistas para recoger información sobre las variables de interés. Los datos recogidos se utilizan para describir las características de una cierta población. Las encuestas se emplean para conocer las actitudes, las creencias, los valores, las características demográficas, los comportamientos, las opiniones, los hábitos, los deseos y las ideas de las personas; además de conseguir otros tipos de información. Se emplean frecuentemente en los negocios, la Administración, la política, la sociología, la sanidad pública, la psicología y la educación debido a la exactitud de la información, obtenida para un gran número de personas a partir de una muestra reducida.

La mayoría de las encuestas describen la incidencia, la frecuencia y la distribución de las características de una población identificada. Además de ser descriptivas, las encuestas pueden usarse también para explorar relaciones entre variables o de una forma explicativa. A continuación se dan ejemplos de temas en cada una de estas tres categorías.

Descriptiva

¿Cuál es la media de tiempo que utilizan los profesores para preparar las lecciones?

Describir las actitudes hacia las ciencias de los alumnos de cuarto curso.

¿Cuáles son las técnicas de asesoramiento más habituales que usan los orientadores de los institutos?

¿Qué piensan los directores sobre los niños de integración que sufren perturbaciones emocionales?

Explorar una relación

¿Los profesores que están a favor del contrato indefinido prueban experiencias de innovación educativa menos a menudo que aquéllos que no están a favor?

¿Existe una relación entre las actitudes de los profesores hacia la disciplina y la satisfacción del estudiante con la clase?

¿Tienen más conflictos las parejas que culpan de los problemas al otro que los que se culpan a sí mismos?

De una forma explicativa

¿Por qué unos directores envían cartas regularmente a los padres y otros raramente lo hacen?

¿Cuál es la razón de que algunos colegios integrados sean verdaderamente integrados mientras otros permanecen segregados?

¿Por qué los estudiantes de un colegio progresan más que estudiantes similares de otro colegio?

La investigación mediante encuesta es muy frecuente en la educación, al igual que en sociología y en los estudios políticos. Hay tres importantes razones para esta popularidad: versatilidad, eficiencia y generalizabilidad (Schutt, 1996). Las encuestas son versátiles porque pueden aplicarse para investigar casi cualquier problema o cuestión. Muchas tesis doctorales recurren a encuestas; los departamentos de estado de educación usan encuestas para determinar los niveles de conocimiento y para averiguar las necesidades para planificar los programas; las escuelas utilizan encuestas para evaluar aspectos del currículo o procedimientos administrativos; las agencias gubernamentales emplean encuestas para formular políticas públicas; las universidades utilizan encuestas para evaluar sus cursos y programas. Muchos de los empleos de las encuestas tienen finalidades prácticas y podrían clasificarse como investigación evaluativa o acción, aunque las encuestas también se utilicen para investigación básica y aplicada.

Las encuestas son habituales porque puede recogerse información creíble a un coste relativamente bajo, especialmente, las encuestas escritas que se envían por correo o las que se distribuyen y recogen de otra manera (por ejemplo, internet es muy barato para hacer encuestas; pero, a menudo las tasas de respuesta no son muy altas). Sobre una base comparativa, por cada euro gastado en una encuesta postal se necesitan cinco euros para la misma encuesta telefónica y diez euros para una encuesta personal. En Estados Unidos, la eficiencia de correos y la casi omnipresencia del servicio telefónico en las familias ayuda a contactar con los individuos rápidamente y a bajo coste. Asimismo, las encuestas son eficaces ya que pueden recogerse datos sobre muchas variables sin que se aumente sustancialmente ni el tiempo ni el coste.

La razón más importante para la reputación de las encuestas es que se pueden seleccionar muestras pequeñas a partir de una población numerosa, de forma que permita generalizaciones hacia esa población. De hecho, las encuestas son, a menudo, el único medio para poder conseguir una descripción representativa de los rasgos, opiniones, actitudes y otras características de la población. Las encuestas también permiten la generalización al resto de la población, en la que pueden compararse subgrupos o contextos diferentes.

Debido al empleo tan frecuente de las encuestas y a que son adaptables a un amplio rango de usos, algunas personas desarrollan la opinión equivocada de que las encuestas son fáciles de realizar. Sin embargo, implican mucho más que simplemente enviar por correo un conjunto de preguntas a un grupo de sujetos. Sin un desarrollo cuidadoso de las preguntas, de los procedimientos de muestreo y del diseño global de la encuesta, no es probable que los métodos de la investigación mediante encuesta proporcionen resultados creíbles. Por ello, es necesario emplear procedimientos de muestreo correctos y diseñar cuidadosamente las técnicas de recogida de datos para asegurar la fiabilidad y la validez.

En la realización de una investigación mediante encuesta resulta de gran ayuda seguir una secuencia de pasos. La secuencia es la que sigue.

1. **Definir la finalidad y los objetivos.** El primer paso es definir el propósito y los objetivos de la investigación. Esto debería incluir un enunciado general y los objetivos generales que definen, con detalle, la información que necesita recabarse. Los objetivos deberían ser lo más claros, cortos y concisos que sea posible. Esto no siempre es tan fácil como parece. Un objetivo, por ejemplo: «el propósito de esta investigación es determinar los valores de los estudiantes universitarios» resulta demasiado vago. ¿Qué se entiende por «valores»? ¿Qué estudiantes universitarios están incluidos? Otra forma de evaluar los objetivos es preguntar si se pretende un empleo concreto de los resultados. A menudo los datos se recogen con la idea de que «estaría bien saber» o «veamos los resultados y luego decidiremos cómo usar los datos». Cualquiera de estas razones para recoger datos es débil. El investigador necesita saber exactamente cómo serán utilizados los resultados antes de que se recojan los datos. Una cuidadosa consideración de los objetivos también ayuda a concretar qué aspectos deben destacarse y cuáles tratarse de una forma más superficial.

2. **Seleccionar los recursos y la población objetivo.** Es necesario tomar decisiones sobre la cantidad total de tiempo, de dinero y de personal disponible antes de diseñar un método específico para la recogida de datos. Lo mejor, por ejemplo, podría ser un cuestionario desarrollado para la ocasión; pero, las limitaciones económicas podrían hacer necesario el uso de una prueba que posea fiabilidad y validez. La cantidad de dinero disponible, también, afectará al tamaño de la muestra que puede extraerse. Los objetivos del estudio también pueden necesitar modificarse debido a las limitaciones financieras. Es preferible hacer bien un estudio pequeño que mal un estudio grande.

También es importante definir la población o el grupo sobre el que el investigador intenta generalizar. La definición es una lista de características o límites para esclarecer la validez externa y especificar el grupo del que se van a recoger los datos. La definición debería considerar las normas del diseño de investigación y el sentido práctico. Si la población es demasiado amplia, como «todos los profesores» o «todos los estudiantes», resultará difícil y costoso obtener una muestra representativa. Si la población es demasiado limitada, como «profesores de educación especial con experiencia de entre cinco y diez años», entonces, la validez externa del estudio será débil.

3. **Elegir y desarrollar técnicas para la recogida de datos.** El cuestionario, la entrevista telefónica y la entrevista personal, descritos en el capítulo 6, son las técnicas empleadas más frecuentemente para la recogida de datos. Estas técnicas deben ser estandarizadas cuando se utilizan en una encuesta para que la información de cada entrevistado se recoja de la misma manera.

4. **Instrucciones.** Es importante especificar instrucciones claras para los entrevistados ya se trate de una entrevista oral, por teléfono o en persona, o de una encuesta escrita u oral. No puede haber ambigüedad sobre cómo, dónde y cuándo deben responder los participantes. Por ejemplo, ¿deben los entrevistados marcar una

casilla, rodear la respuesta correcta o hacer marcas sobre un espacio? Si hay una hoja diferente para las respuestas (lo que generalmente no es recomendable; si intenta usar un formato de lectura óptica, lo mejor es tener las preguntas impresas en el formulario), ser claro sobre los «tachones», mediante el empleo de lápiz, si es necesario (algunos formularios de lectura óptica sólo necesitan un área oscurecida, que no tiene que ser necesariamente de lápiz) y administrar la encuesta en persona en lugar de por correo. Además, las instrucciones deberían indicar de forma clara cómo se deben devolver los cuestionarios. Esto puede incorporarse en la carta de envío, al principio de la encuesta o al final de ésta. La mejor política es incluir las instrucciones para completar y devolver la encuesta sobre el mismo papel enviado.

5. **Muestreo.** Los principios básicos del muestreo se presentaron en el capítulo 4. La mayoría de las encuestas usan el muestreo probabilístico para asegurar una representación adecuada de la población. El muestreo aleatorio se estratifica a menudo sobre algunas variables, como por ejemplo, sexo, curso, capacidad, nivel y situación socioeconómica.

6. **Carta de difusión.** En el caso de los cuestionarios por correo, la forma del documento adjunto o de la carta de difusión es crucial para determinar el porcentaje de sujetos que devolverán el formulario completo. La carta debería ser breve y establecer la credibilidad del investigador y del estudio. Esto se consigue incluyendo los nombres e identificaciones de los investigadores, la finalidad y la intención del estudio a grandes rasgos, la importancia del estudio para el entrevistado, la importancia del entrevistado para el estudio, la protección que se ofrece al entrevistado de que sus datos son confidenciales, un tiempo límite para devolver la encuesta cumplimentada, que no será ni muy corto ni muy largo (normalmente una semana, más o menos), los apoyos al estudio de instituciones o grupos reconocidos, una breve descripción del cuestionario y del procedimiento, una mención a la posibilidad de obtener los resultados, una petición de cooperación y honradez y un agradecimiento al entrevistado. Naturalmente, es más probable que el entrevistado coopere si la carta es ordenada, con apariencia profesional y sin errores de ortografía o de gramática. Es mejor si la carta informa de que la investigación está dirigida por una organización fiable para el entrevistado. Si es posible, las cartas de difusión deberían ser personalizadas. El extracto 7.9 contiene una carta ejemplo que contiene todos estos puntos.

7. **Prueba piloto.** Es indispensable realizar una prueba piloto tanto de las instrucciones como de la encuesta antes de la distribución de la muestra identificada. Esta prueba piloto debería ser llevada a cabo con entrevistados similares a los que compondrán la muestra. Generalmente, el estudio piloto puede identificar las necesidades de cambio si al menos diez individuos están dispuestos a completarlo y a proporcionar sugerencias para mejorar la claridad y el formato. La prueba piloto también nos proporciona una estimación de la cantidad de tiempo que llevará completar la encuesta. Finalmente, la prueba piloto proporciona una idea

inicial del patrón de respuestas más probable y si se necesitan revisiones para evitar los efectos de techo o suelo. Ocasionalmente, se realiza una prueba piloto inicial, seguida por una prueba piloto más formal sobre una muestra mayor.

8. **Seguimiento.** El envío inicial del cuestionario, la carta adjunta y el sobre franqueado para la respuesta tendrá como resultado un índice de respuesta de entre el 40 y el 60 por ciento, esto es, entre el 40 y el 60 por ciento de la muestra devolverá los cuestionarios. Este índice es más alto para cuestionarios cortos con poca importancia y menor para cuestionarios largos sobre temas personales. El procedimiento tipo para el seguimiento es primero, el envío de una tarjeta recordatorio entre los diez días y las dos semanas posteriores al envío inicial. Para los que no han contestado, se manda una carta de seguimiento tres o cuatro semanas después, que incluirá una copia de la encuesta. Si el índice de respuesta todavía sigue siendo bajo y tenemos presupuesto, se puede mandar una carta certificada con otra copia de la encuesta a las seis u ocho semanas. En el caso de las encuestas que aseguraban el anonimato, es imposible saber quién ha contestado el cuestionario. En este caso los investigadores pueden recurrir a varios procedimientos: mandar una carta de seguimiento a todo el mundo, indicando que quien haya devuelto el cuestionario puede ignorarla; también el investigador puede codificar los cuestionarios sin el conocimiento de los sujetos para identificar a los entrevistados; o bien los sujetos pueden enviar una tarjeta independiente cuando devuelven el cuestionario. La tarjeta de seguimiento normalmente aporta entre un 10 y un 30 por ciento más de respuestas y un segundo seguimiento aumenta este índice entre un 10 y un 20 por ciento. Si un investigador puede obtener un índice total del 70% o más, entonces lo está haciendo muy bien.

EXTRACTO 7.9

CARTA DE DIFUSIÓN

Estimado miembro del consejo de dirección de la Asociación de Padres,

Su actitud ante la educación es importante.

Importancia del entrevistado

Su compromiso con su hijo y con su educación está demostrado por su puesto en la Asociación de Padres. Como miembro del consejo de dirección es usted una persona esencial en muchos aspectos. Usted escucha a los padres y habla con ellos. Escucha y habla con el director y trabaja con el/ ella. Usted es un vínculo muy importante entre el colegio y la comunidad. Sus decisiones tienen una gran importancia para sus hijos y su escuela.

Importancia del estudio

El puesto de los directores de primaria también es importante. Usted conoce muy bien las responsabilidades que ellos tienen y que sus decisiones afectan diariamente a sus hijos y pueden afectarles para el resto de sus vidas.

Identificación del investigador; propósito del estudio; importancia del entrevistado	{	<p>Estoy realizando un estudio en la universidad de Saint John de Nueva York que está dirigido a los miembros de la Asociación de Padres del consejo de dirección y a los directores de primaria. Como un miembro del consejo de dirección, usted y el cargo que desempeña son de particular interés para este estudio.</p> <p>Este es un estudio nacional. Usted y todos los participantes han sido seleccionados de forma aleatoria. Se ha solicitado permiso al delegado para pedir su colaboración.</p>
Pedir colaboración	{	<p>Compartimos un interés y una preocupación común sobre la investigación. Sobre esta base, es una meta común el incrementar el conocimiento sobre la educación por lo que solicito su colaboración rellenando el cuestionario adjunto.</p>
Descripción del cuestionario	{	<p>El cuestionario contiene tres partes. Las dos primeras tratan de sus impresiones sobre el director y la tercera solicita información personal sobre usted.</p>
Respaldo del estudio	{	<p>El cuestionario ha sido estudiado y evaluado de forma cuidadosa por profesionales y miembros del consejo de dirección de Asociaciones de Padres.</p> <p>Pedimos a miembros de consejos de dirección de Asociaciones de Padres que rellenaran el cuestionario. De forma unánime dijeron que no se tardaba más de veinte minutos en cumplimentarlo.</p>
Protección ofrecida al entrevistado	{	<p>No tiene que firmar el cuestionario y puedo asegurarle que sus respuestas serán anónimas y confidenciales. Su participación por supuesto es voluntaria.</p>
Descripción del procedimiento (tiempo límite)	{	<p>Por favor, conteste las preguntas y devuelva el cuestionario en sobre cerrado lo antes posible</p>
Oportunidad de obtener resultados	{	<p>Si desea un resumen del estudio, marque la casilla correspondiente al final del cuestionario y envíe su nombre y su dirección en la tarjeta adjunta.</p>
Agradecimiento al entrevistado	{	<p>Gracias por su colaboración.</p>

Atentamente
 Stanley Cogan
 Asistente del director
 Escuelas públicas de Nueva York

Fuente: «The relationship of the social class and personal characteristics of parent association executive board members to the role expectations and personal characteristics which they advocate in the selection of elementary principals in selected American cities», conferencia doctoral de Stanley Cogan. Reproducido con permiso de Stanley Cogan.

9. *No contestadas*. En la mayoría de las encuestas habrá un porcentaje de sujetos que no devuelvan el cuestionario cumplimentado. A estos sujetos se les denomina «sin respuesta» y el investigador necesitará hacer esfuerzos adicionales para comprobar si la inclusión de estos sujetos podría alterar los resultados. En la mayoría de las encuestas por correo con una muestra grande (200 o más), las no contestadas, probablemente, no afectarán a los resultados de una forma apreciable si el índice de participación es por lo menos del 70%. Si los resultados van a emplearse para adoptar decisiones importantes, o si la naturaleza de las preguntas podría provocar que algunos tipos de sujetos no contestasen, entonces, las no contestadas deben ser tenidas en cuenta. Una aproximación para la investigación de la posibilidad de un grupo sesgado de entrevistados es obtener una muestra aleatoria de los no entrevistados. Si es posible, se entrevista a estos individuos y se comparan sus respuestas con las de los que sí contestaron el cuestionario. Si las respuestas son semejantes, el investigador puede asegurar que los cuestionarios obtenidos representan una muestra sin sesgo. Si es imposible entrevistar a los no entrevistados seleccionados de forma aleatoria, el siguiente paso es compararlos con los sujetos que respondieron con respecto a características demográficas. Si las respuestas o las características demográficas de los no entrevistados son diferentes, entonces la diferencia debería aparecer en la exposición e interpretación de los resultados del estudio. Debería prestarse especial atención a los estudios con un índice de respuesta relativamente bajo (menor del 70%), y sin un análisis de la forma de los no entrevistados, pueden cambiar los resultados. La investigación ha demostrado que el índice de respuesta para las encuestas por correo puede incrementarse sustancialmente mediante el uso de gratificaciones económicas (Hopkins y Gullickson, 1992). Una gratificación de 1 euro puede incrementar el índice de respuesta una media del 20%.

Antes de abandonar el tema de las entrevistas, queremos volver a destacar que la investigación mediante encuesta no es sencilla de preparar y de dirigir. Intervienen muchos factores relacionados con el formato de los ítems, el enfoque de las preguntas, la redacción y otras variables que deben ser tenidos en cuenta. Para una excelente revisión de estos problemas, ver Schuman y Presser (1996) y para información adicional sobre el desarrollo y dirección de encuestas, *The Survey Kit*².

Debido a que la mayoría de las encuestas utilizan ítems de formato cerrado que son analizados por ordenador, es esencial para los investigadores consultar con expertos que pueden ayudarles en el diseño de la encuesta para una entrada de datos exacta y fácil. Tales expertos también pueden colaborar en la revisión de la redacción y otros aspectos técnicos de la encuesta.

² *The Survey Kit* es una serie de volúmenes publicados por Sage Publications, Inc que recoge cuestiones de desarrollo, realización, análisis y presentación de datos de encuestas.

INVESTIGACIÓN *EX POST FACTO*

En los diseños descriptivos y correlacionales, en la mayoría de los casos, no se estudian las relaciones causales. Sin embargo, existen ciertos diseños no experimentales que se usan para investigar relaciones causales. Uno de éstos se denomina *ex post facto*. En estos estudios, la finalidad de la investigación es examinar cómo una variable independiente identificada produce cambios en la variable dependiente, pero las circunstancias no permiten un diseño experimental. Considere la siguiente lista de preguntas de investigación. En cada caso, la relación causa-efecto implicada descarta una manipulación experimental de la variable independiente.

¿Cuál es el efecto de la asistencia a una guardería sobre el grado de sociabilidad de los niños?

¿Cuál es el efecto de las familias monoparentales en el rendimiento?

¿Los profesores que se graduaron en universidades de artes liberales tienen mayor permanencia en la enseñanza que los que se graduaron en escuelas de magisterio?

¿Cuál es la relación entre la participación en actividades extraescolares y la autoestima?

Características de la investigación *ex post facto*

Es simplemente imposible, poco ético o poco factible manipular variables como por ejemplo, ser o no familia monoparental, asistencia a una guardería o la elección de la universidad por parte de los estudiantes, así como muchas otras variables como raza, nivel socioeconómico y personalidad. Un investigador, probablemente, tendría alguna dificultad en la asignación aleatoria de los niños tanto si asisten como si no lo hacen a una guardería.

Aunque es deseable estudiar las relaciones de causa y efecto en tales situaciones, las circunstancias de la investigación hacen que no pueda llevarse a cabo esta manipulación de la variable independiente. El tipo de diseño que se emplea en la mayoría de estas situaciones se llama *ex post facto*, o comparativo causal. El propósito de la **investigación *ex post facto*** es averiguar si una o más condiciones previas han causado las posteriores diferencias entre los grupos de sujetos. En otras palabras, el investigador identifica las condiciones que ya han tenido lugar (*ex post facto* es «después del hecho» en latín), y entonces recoge datos para investigar la relación de estas circunstancias sobre el comportamiento posterior. En la investigación *ex post facto*, el investigador intenta determinar si las disparidades entre los grupos (la variable independiente) ha dado lugar a una diferencia observada en la variable dependiente.

Los diseños *ex post facto* se confunden fácilmente con los diseños experimentales ya que ambos tienen una finalidad semejante (determinar las relaciones

causa-efecto), las comparaciones entre grupos y el uso de análisis estadísticos, y vocabulario similares para la descripción de los resultados. Sin embargo, en los estudios experimentales y semiexperimentales, el investigador controla, de forma deliberada, el efecto de alguna condición por medio de la manipulación de la variable independiente, mientras que en la investigación *ex post facto* no existe manipulación de las condiciones, ya que la posible causa ocurrió antes de que se iniciara el estudio. En los diseños *ex post facto*, por lo tanto, normalmente hay un tratamiento y un grupo de control, un factor que puede fomentar la confusión de la investigación con aproximaciones experimentales.

Los diseños *ex post facto* también se confunden con la investigación correlacional y comparativa ya que los tres implican la no manipulación y poseen limitaciones similares para la interpretación de los resultados. Sin embargo, los diseños *ex post facto* intentan identificar relaciones causales, mientras que la investigación comparativa y correlacional no lo hacen.

Realización de la investigación *ex post facto*

Aunque en la investigación *ex post facto* no puede manipularse la variable independiente y es imposible una asignación al azar de los sujetos a los grupos, pueden usarse varios procedimientos para la planificación de la investigación *ex post facto* que aumenten el control y limiten las probables hipótesis rivales. El primer paso consiste en la formulación de un problema de investigación que incluya las posibles causas de la variable dependiente. La elección de las posibles causas se basa en investigaciones previas y en la interpretación que los investigadores dan a las observaciones de los fenómenos que se están estudiando. Suponga, por ejemplo, que el investigador quiere investigar el efecto del tamaño de la clase en el rendimiento y es imposible asignar estudiantes de forma aleatoria a diferentes tamaños de clase para llevar a cabo un experimento real. El interés del investigador puede basarse en una investigación correlacional que muestre una relación negativa entre el tamaño de la clase y el rendimiento, y las observaciones sobre los alumnos de clases más reducidas no parecen ser mejores. Luego, el problema de investigación es: ¿cuál es la repercusión del tamaño de la clase en el rendimiento?

El segundo paso es identificar las posibles hipótesis alternativas que podrían explicar la relación. Por ejemplo, el investigador podría enumerar como posibles causas de un mejor rendimiento en las clases reducidas varios factores, por ejemplo: las clases poco numerosas poseen mejores profesores; los estudiantes con una mayor capacidad están en las clases más reducidas; los estudiantes con mayor motivación están en las clases más reducidas; a las clases menos numerosas asisten más estudiantes de ambiente socioeconómico alto que estudiantes con un entorno socioeconómico bajo; y quizá, los profesores de las clases más reducidas emplean un tipo de enseñanza diferente al que utilizan los profesores de las clases más nume-

rosas. Cada uno de estos factores podría relacionarse con la causa de que los estudiantes de las clases más reducidas rindan más que los de las clases numerosas.

El tercer paso es encontrar y seleccionar los grupos que serán comparados. En nuestro ejemplo sobre el tamaño de la clase y rendimiento, el investigador necesitará, inicialmente, definir de forma matemática tanto el tamaño de la clase como del rendimiento. Una clase reducida podría tener menos de quince alumnos y una clase numerosa más de veinticinco. El rendimiento puede definirse como la adquisición de conocimientos por parte del estudiante mientras está en la clase. El investigador también necesita identificar los cursos y su localización. Supongamos en este caso que el investigador está interesado en los niveles de primaria y la población accesible pertenece a un único distrito escolar. Una vez que se han definido las variables, los grupos deben seleccionarse de manera que sean lo más homogéneos posible en lo que se refiere a las características que constituyen la hipótesis alternativa y que sean diferentes con respecto a la variable independiente. En nuestro ejemplo, el investigador selecciona grupos que se diferencian en el tamaño de la clase pero que son similares en factores como capacidad, ambiente socioeconómico, métodos de enseñanza, calidad de los profesores y motivación de los estudiantes. El emparejamiento es una buena aproximación para la formación de grupos que serán lo más homogéneos posible con respecto a los factores que afectan a la variable dependiente. Por ejemplo, en nuestro estudio del tamaño de clase, podría emparejar y seleccionar los alumnos sobre la base de la capacidad inicial, de forma que sólo estarían incluidos en el análisis los estudiantes con el mismo nivel de capacidad, aunque existan otros estudiantes en las clases.

El cuarto paso es recoger y analizar los datos sobre los sujetos, incluyendo datos sobre factores que pueden constituir hipótesis alternativas. Debido a que la investigación *ex post facto* es posterior al hecho, la mayoría de los datos necesarios ya han sido recabados y sólo necesitaríamos recoger los datos de las fuentes apropiadas. El análisis de datos es muy similar a los procedimientos utilizados en estudios experimentales o semiexperimentales en los que los grupos se comparan según las variables de interés. Por ejemplo, en nuestro estudio, puede obtenerse la media de todas las puntuaciones de rendimiento y compararse con la media del rendimiento en las clases numerosas. Los datos de las variables externas también se comparan y se incorporan a los análisis estadísticos para ayudar a establecer juicios sobre hipótesis alternativas verosímiles.

En la interpretación de los resultados de una investigación *ex post facto*, debe destacarse que los enunciados causa-efecto únicamente deben hacerse con precaución. En nuestro ejemplo, si se encuentran diferencias en el rendimiento entre los grupos, el investigador puede entonces concluir que existe una relación entre el tamaño de la clase y el rendimiento. Los resultados no significan, inequívocamente, que asistir a una clase más numerosa o más reducida tenga un efecto causal en el rendimiento. Puede haber una posible relación causa-efecto,

pero esto depende de la habilidad del investigador para seleccionar grupos de comparación homogéneos en todas las variables importantes excepto el tamaño de la clase y en la confianza con la que podamos descartar otras posibles hipótesis alternativas. Por ejemplo, si resultó que todas las clases reducidas procedían de una escuela y las clases numerosas de otra, entonces las políticas o métodos específicos de una escuela y no relacionados con el tamaño de la clase pueden constituir posibles hipótesis alternativas.

Un buen ejemplo de investigación *ex post facto* es un estudio que investiga las consecuencias sobre el juego en las clases de preescolar con alumnos de la misma edad o de distintas edades. Como se indica en el extracto 7.10, se compararon las modalidades de juegos y la interacción social de los niños en cuatro clases de la misma edad con clases a las que asistían niños de varias edades. Observe que los investigadores resumen, en el segundo párrafo del extracto, cómo las clases de la misma edad y de edades distintas eran comparables con respecto al esquema físico, materiales, ratio profesor-alumnos, factores socioeconómicos de los niños y la cantidad de tiempo que los niños permanecían en clase. Probablemente estas características se reconocieron antes de recogerse los datos para descartar estas posibles hipótesis alternativas. Por supuesto pueden darse otras diferencias entre los grupos, además de la variable edad, que podrían afectar a los resultados. Por ejemplo, una disparidad entre los monitores podría afectar al juego y a la interacción social.

EXTRACTO 7.10

INVESTIGACIÓN EX POST FACTO

MÉTODO

Sujetos. Se observaron las modalidades sociales y cognitivas de juego y las respuestas de los compañeros a éstas en dos clases de niños de 3 años (N = 40), dos clases de niños de 4 años (N = 32) y dos clases de niños con edades mixtas de 3 y 4 años (N = 36). La clase era considerada mixta si al menos el 40% de los niños tenía 3 años y el 40% tenía 4 años. Los niños se consideraban de 3 años en el rango que iba desde los 2 años y 9 meses hasta los 3 años y 8 meses; y de 4 años en el rango de 3 años y 9 meses hasta 4 años y 8 meses cuando comenzó el estudio. Para conceder las mismas oportunidades a las actividades del mismo sexo y a las mixtas, se seleccionaron clases en las que al menos el 40% fueran niños y el 40% niñas. Estos criterios para la clasificación de las clases eran similares a los señalados por la investigación previa (Goldman, 1981; Roopnarine, 1984). Se utilizaron estos criterios porque queríamos estudiar grupos de edad semejantes para hacer comparaciones con los resultados de los estudios previos (los de Goldman).

Todas las clases tenían una orientación centrada en los niños y se compararon con respecto al esquema físico, materiales y ratio profesor-alumno (1:9). Los alumnos procedían de entornos socioeconómicos medios, valorados a través del índice de cuatro factores

de condición social de Hollingshead (Hollingshead, sin fecha). Los análisis preliminares revelaban que no existían diferencias significativas en los factores sociodemográficos entre las clases de la misma edad y las de edades distintas.

Además, los niños de ambos tipos de clase estaban matriculados con aproximadamente la misma antelación en las clases de distinta edad antes del estudio (clases de niños de la misma edad $x = 6$ meses; clases de niños de distinta edad, $x = 5$ meses).

Fuente: «Social-Cognitive Play Patterns in Same-Age and Mixed-Age Preschool Classrooms», de N. S. Mounts y J. L. Roopnarine, *American Educational Research Association*, Washington D. C. Reimpreso con permiso.

NORMAS DE ADECUACIÓN

Al juzgar la adecuación de la investigación descriptiva, comparativa, correlacional, de encuesta y *ex post facto* nos será de gran ayuda tener en cuenta las siguientes preguntas. Están organizadas con el objeto de centrar su atención en los criterios más importantes para el diseño y la evaluación de estos tipos de investigación.

Investigación descriptiva

1. ¿El problema de investigación es claramente descriptivo por naturaleza?
2. ¿Hay una descripción precisa de la muestra, de la población y de los procedimientos para el muestreo?
3. ¿Proporcionará la muestra resultados sesgados o distorsionados?
4. ¿Son válidas y fiables las puntuaciones resultantes de los instrumentos de medida?
5. ¿Las presentaciones gráficas de los resultados distorsionan los hallazgos?
6. ¿Se hacen relaciones inapropiadas o conclusiones causales a partir de los resultados descriptivos?
7. Si es un estudio transversal, ¿afectan las diferencias de los sujetos a los resultados?
8. Si es longitudinal, ¿la pérdida de sujetos es una limitación?
9. ¿Se emplean las diferencias entre grupos para identificar posibles relaciones?

Investigación comparativa

1. ¿Indica, de forma evidente, el problema de investigación que las relaciones serán investigadas a través de la comparación de las diferencias entre grupos?
2. ¿Es clara la descripción de sujetos, del muestreo y de las pruebas realizadas?
3. ¿Sesgan los resultados los procedimientos de muestreo?
4. ¿Hay suficiente evidencia de fiabilidad y validez?
5. ¿Se hacen deducciones causales inapropiadas?

Investigación correlacional

1. ¿Indica claramente el problema de investigación que se estudiarán relaciones?
2. ¿Hay una descripción rigurosa del muestreo empleado? ¿Proporcionará la muestra suficiente variedad de respuestas para obtener una correlación?
3. ¿Son válidas y fiables las puntuaciones que resultan de la instrumentación?
4. ¿Hay un rango restringido de puntuaciones?
5. ¿Hay factores que podrían contribuir al hallazgo de correlaciones espúreas?
6. ¿Se emplea la aproximación «de escopeta» en el estudio?
7. ¿Se hacen deducciones causales inapropiadas a partir de los resultados?
8. ¿Qué tamaño tiene la muestra? ¿Podría el tamaño de la muestra afectar al «significado» de los resultados?
9. ¿Se confunde el coeficiente de correlación con el coeficiente de determinación?
10. Si se han hecho pronósticos, ¿están basados en una muestra diferente?
11. ¿Es el tamaño de la correlación lo suficientemente grande para las conclusiones obtenidas?

Investigación mediante encuesta

1. ¿Están claros los objetivos y propósitos de la encuesta?
2. ¿Es probable que la población objetivo y los procedimientos de muestreo proporcionen una respuesta creíble a la(s) pregunta(s) de investigación?
3. ¿Está el conjunto de pruebas correctamente diseñado y redactado? ¿Ha pasado una prueba piloto? ¿Es apropiado a las características de la muestra?
4. ¿Se asegura la confidencialidad de las respuestas? Si no es así, ¿es probable que este hecho afecte a los resultados?
5. ¿Establece la carta de difusión la credibilidad de la investigación? ¿Hay alguna posibilidad de que lo que se exponga en la carta pueda sesgar las respuestas?
6. ¿Cuál es el índice de devolución? Si es bajo o está en el límite, ¿ha habido algún seguimiento para las no contestadas?
7. ¿Reflejan las conclusiones el índice de devolución y otras posibles limitaciones?

Investigación ex post facto

1. ¿Era la investigación de las relaciones causa-efecto el objetivo principal del estudio?
2. ¿Han ocurrido ya las presuntas condiciones de causa-efecto?
3. ¿Hubo manipulación de la variable independiente?

4. Cuándo se compararon los grupos, ¿eran ya diferentes con respecto a la variable independiente?
5. ¿Se reconocieron y consideraron como probables hipótesis alternativas las variables externas?
6. ¿Se han enunciado afirmaciones causales con respecto a los resultados con la suficiente prudencia?
7. ¿Se han mencionado referencias a la validez externa en las conclusiones?

RESUMEN

Este capítulo ha proporcionado una revisión de la investigación descriptiva, correlacional, mediante encuesta y *ex post facto*, subrayando las normas de diseño que afectan a la calidad de la investigación. Los principales puntos del capítulo son:

1. La investigación descriptiva se preocupa de la situación actual o pasada de algo.
2. Los estudios evolutivos investigan los cambios de los sujetos a través del tiempo y son longitudinales y transversales.
3. Los estudios de relación comparativos examinan las diferencias entre grupos.
4. Los estudios de relación correlacionales manejan, al menos, dos puntuaciones obtenidas de cada sujeto.
5. Los estudios comparativos y correlacionales deberían definir de forma rigurosa el muestreo, la instrumentación y los procedimientos para la recogida de datos.
6. La selección de los sujetos y de los instrumentos en la investigación correlacional debería asegurar un rango de respuestas para cada variable.
7. En la investigación predictiva, la variable criterio se predice a partir de un comportamiento previo medido por la variable independiente.
8. La correlación múltiple permite la combinación de varias variables independientes en relación con la variable dependiente o la variable criterio.
9. La correlación nunca debería inferir causalidad por terceras variables que no han sido medidas y a su incapacidad para evaluar dirección causal entre las dos variables.
10. Las correlaciones espúreas sobreestiman o infravaloran la relación real entre dos variables.
11. El coeficiente de determinación es el cuadrado del coeficiente de correlación y estima la varianza compartida por las variables.

12. Los coeficientes de correlación deberán interpretarse cuidadosamente cuando son estadísticamente significativos con bajas relaciones y un gran número de sujetos.
13. Las decisiones referidos a individuos requieren mayores correlaciones que para grupos.
14. La investigación mediante encuesta utiliza cuestionarios para describir las características de las poblaciones.
15. Las encuestas se utilizan con propósitos descriptivos, de relación o explicativos.
16. En las encuestas por correo, la naturaleza de la carta de difusión es crucial para la obtención de un índice aceptable de respuesta del 60% o mayor.
17. Los diseños *ex post facto* se emplean para estudiar las posibles relaciones causales después de que la causa potencial haya tenido lugar.
18. En la investigación *ex post facto* se seleccionan los sujetos sobre la base de los grupos en los que estaban en un momento dado; probablemente no haya asignación aleatoria de sujetos a grupos y no hay manipulación activa de la variable independiente.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La investigación descriptiva estudia...
 - a. cuestiones causales y de diferencia.
 - b. variables dependientes como el rendimiento.
 - c. el estado actual de los asuntos.
 - d. relaciones de causa-efecto.
2. Los estudios longitudinales investigan...
 - a. sujetos diferentes a través del tiempo.
 - b. relaciones transversales.
 - c. sujetos diferentes en el mismo momento.
 - d. los «mismos» sujetos a través del tiempo.
3. El rango restringido y las muestras homogéneas son especialmente problemáticos para...
 - a. los estudios de escopeta.
 - b. los estudios de relación.
 - c. los estudios descriptivos.
 - d. los estudios longitudinales.
4. En un estudio de predicción la medición de la variable independiente...
 - a. tiene lugar antes de la medida de la variable dependiente.

- b.** tiene lugar al mismo tiempo que la de la variable dependiente.
 - c.** debería hacerla el investigador directamente.
 - d.** tiene lugar después de la medida de la variable dependiente.
- 5.** Los estudios de correlación múltiple...
 - a.** comparan los pesos beta para los coeficientes de regresión.
 - b.** manejan muchas variables independientes.
 - c.** manejan muchas variables dependientes.
 - d.** comparan R con los pesos beta.
- 6.** Las correlaciones son espúreas si...
 - a.** son demasiado altas.
 - b.** se interpretan de forma incorrecta.
 - c.** son demasiado bajas.
 - d.** se sobreestima o se infravalora la verdadera relación.
- 7.** Las correlaciones no deberían emplearse para deducir causalidad porque...
 - a.** sólo se estudian dos variables.
 - b.** no están bien controladas.
 - c.** hay terceras variables no medidas.
 - d.** sólo puede usarse la estadística inferencial para deducir causalidad.
- 8.** La investigación mediante encuesta emplea las siguientes técnicas:
 - a.** entrevistas.
 - b.** observación.
 - c.** cuestionarios.
 - d.** la a y la c son correctas.
 - e.** todas son correctas.
- 9.** Las encuestas pueden utilizarse para cada uno de los siguientes propósitos excepto...
 - a.** causal.
 - b.** de relación.
 - c.** explicativo.
 - d.** descriptivo.
- 10.** ¿Cuál de las secuencias describe el orden correcto de pasos para la realización de una investigación mediante encuesta?
 - a.** definir el propósito, identificar las técnicas, identificar la población, muestreo, correo, seguimiento.
 - b.** definir el propósito, muestreo, identificar las técnicas, correo, seguimiento.
 - c.** definir el propósito, identificar la población, muestreo, identificar las técnicas, correo, seguimiento.
 - d.** definir el propósito, identificar la población, identificar las técnicas, muestreo, correo, seguimiento.

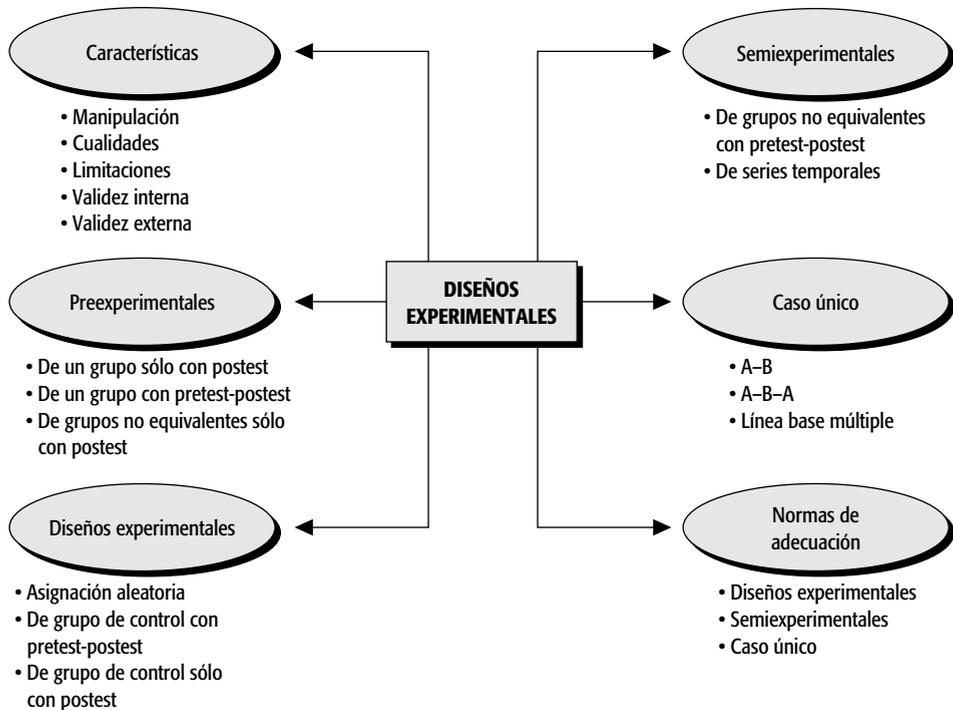
11. Dos características de la investigación *ex post facto* son...
 - a. asignación aleatoria y con comparación de grupos.
 - b. sin manipulación y sin asignación aleatoria.
 - c. sin asignación aleatoria y con manipulación de la variable independiente.
 - d. sin manipulación y con asignación aleatoria.
12. La selección de grupos de comparación apropiados es crucial en la investigación *ex post facto* debido a que...
 - a. los grupos deberían ser iguales en todos los aspectos excepto en la variable independiente.
 - b. los grupos deberían ser aproximadamente iguales respecto a la variable dependiente.
 - c. de lo contrario los grupos no serían comparables.
 - d. si los grupos se eligen de forma incorrecta, las hipótesis alternativas pueden ser más probables.

Problemas aplicados

1. En un estudio de motivación y aprendizaje, un profesor empleó la siguiente estrategia de investigación: se preguntó a los estudiantes qué estrategias de enseñanza les resultaban más motivadoras. El profesor, luego, consultó las notas previas para investigar si éstas habían sido superiores en las sesiones más motivadoras comparadas con las sesiones menos motivadoras. El profesor encontró que, ciertamente, el rendimiento de los estudiantes subió conforme el profesor utilizaba técnicas progresivamente más motivadoras. A partir de este resultado decidió emplear las técnicas más motivadoras todo el tiempo. ¿Su decisión de emplear las técnicas más motivadoras fue correcta? ¿Por qué sí o por qué no?
2. Indicar a partir de las siguientes descripciones si la investigación es descriptiva, comparativa, correlacional, predictiva o *ex post facto*.
 - a. Un investigador descubre que hay una relación positiva entre las calificaciones de los alumnos y su atención.
 - b. Un investigador descubre que los estudiantes que son más solitarios tienen padres más permisivos que los estudiantes menos solitarios.
 - c. El decano de la Escuela de Magisterio utiliza puntuaciones SAT para identificar a los estudiantes que pueden tener problemas con los exámenes nacionales de profesores.
 - d. Los niños que pertenecen a hogares con más apoyo y afecto tienen una mayor autoestima que los niños de hogares con rechazo.

- e. La mayor parte de las facultades de los Estados Unidos está a favor de acabar con los contratos indefinidos.
- f. Los investigadores han encontrado que los estudiantes graduados que asistieron a una universidad pequeña están más capacitados en la expresión escrita que los estudiantes de una universidad grande.

Diseños de investigación experimentales y de caso único



PALABRAS CLAVE

experimental	diseño de grupo de control sólo con postest
variables asignadas	diseños semiexperimentales
variables atributo	diseño de grupos no equivalentes con pretest-postest
grupo experimental o de tratamiento	diseño de series temporales
grupo de control o de comparación	diseño de series temporales interrumpidas con un grupo
diseños preexperimentales	diseño de series temporales interrumpidas con grupo de control
diseño de caso único sólo con postest	diseños de caso único
diseño de caso único con pretest-postest	línea base
diseño pretest-postest	diseño A-B
diseño de grupos no equivalentes sólo con postest	diseño A-B-A
diseños experimentales puros	diseño de reversión
diseño de grupos de control con pretest-postest	diseños de línea base múltiple

En el capítulo 2 se presentaron los conceptos experimento y modalidad de investigación experimental. En este capítulo consideraremos la realización de los experimentos de modo más detallado. Nos centraremos en las cuatro principales categorías de los diseños experimentales empleadas para hacer deducciones causales sobre la relación entre las variables independientes y dependientes: preexperimentales, experimentales, semiexperimentales y de caso único.

Expondremos el propósito, las cualidades y las limitaciones de la investigación experimental en general, así como de diseños más específicos. Desde el comienzo insistimos en que en cualquier estudio en el que se infiera causalidad, el investigador debe descartar las hipótesis alternativas verosímiles que podrían explicar los resultados.

Los criterios de adecuación para valorar los diseños están resumidos al final del capítulo.

UNA INTRODUCCIÓN A LA INVESTIGACIÓN EXPERIMENTAL

El término *experimento*, como muchos otros términos, puede poseer varios significados. Definido de manera general, los experimentos son simplemente un modo de saber algo variando algunas condiciones y observando el consiguiente

efecto. En otras palabras, cambiamos algo y observamos el efecto. Como humanos, recurrimos constantemente a experimentos naturales para aprender algo: los niños pequeños experimentan con una serie de tácticas para saber cuál afectará más a papá o a mamá; los profesores aplican un nuevo enfoque de la disciplina para ver si funciona; y los alumnos varían las técnicas de estudio para ver con cuál de ellas obtienen mejores notas. Este comportamiento simple de ensayo y error es un intento de demostrar la causalidad, que es el principal objetivo de un experimento. La diferencia entre estos experimentos y los muy sofisticados realizados por científicos de laboratorio es el grado en que el experimentador puede estar seguro de que las condiciones modificadas causaron el efecto observado. Por lo tanto, la interpretación de la causalidad es un elemento esencial en la investigación experimental.

Características de la investigación experimental

Según la definición tradicional, la investigación experimental posee seis características distintivas: equivalencia estadística de sujetos en diferentes grupos, obtenida normalmente mediante asignación aleatoria de sujetos; comparación de dos o más grupos o conjuntos de condiciones; manipulación directa de al menos una variable independiente; medición de cada variable dependiente; empleo de la estadística inferencial, y un diseño que facilite el potencial para el control máximo de las variables extrañas. Estas características están normalmente presentes en la investigación experimental física y biológica pero, como observamos en el capítulo 2, estas condiciones raramente pueden alcanzarse completamente en la investigación educativa. Esto, sin embargo, no resta importancia al método experimental en educación. Muchas investigaciones educativas se aproximan a la mayor parte de estas características y necesitamos entender el modo cómo los diferentes métodos de realización de una investigación, que examina las relaciones causales, afecta a la interpretación de los resultados. Esto es lo que Campbell y Stanley (1963) tenían en mente cuando escribieron su clásico e influyente capítulo «*Experimental and Quasi-Experimental Designs for Education*» (Diseños experimentales y semiexperimentales en educación). La siguiente cita deja clara su visión en aquel momento:

Este capítulo está dedicado al experimento: como el único medio para crear controversias con respecto a las prácticas educativas, como la única manera de verificar las mejoras en educación y como el único modo de establecer una tradición acumulativa en la que los avances puedan introducirse sin el peligro de dejar de lado la vieja sabiduría en beneficio de novedades inferiores (Campbell y Stanley, 1963, p. 2)

Por lo tanto, necesitamos distinguir entre lo que debe ser considerado como investigación **experimental pura**, en la que las seis características anteriores

están totalmente presentes, y el método de investigación experimental más genérico, en el que las características están parcialmente presentes. En este capítulo, dado que nos ocupamos de la investigación en educación, interpretamos el adjetivo experimental en su contexto más general.

La primera característica de la investigación experimental, la obtención de equivalencia estadística de sujetos en diferentes grupos, es necesaria para que muchas importantes y posibles amenazas a la validez interna no sean factores en la interpretación de los resultados. En otras palabras, el investigador quiere hacer que los grupos que serán comparados tengan los mismos rasgos y características de modo que cualquier diferencia existente en los resultados no pueda ser atribuida a diferencias entre los grupos o experiencias que tuvieron lugar durante el experimento. La asignación aleatoria de sujetos es el método más empleado para conseguir la equivalencia estadística de los grupos. También se utilizan con esta finalidad otros métodos como el emparejamiento de sujetos o de grupos o usar sujetos como sus propios controles.

La segunda característica sugiere que tienen que compararse, al menos, dos grupos o condiciones. Un experimento no puede realizarse con un grupo de sujetos en una condición determinada en un momento determinado. El propósito de un experimento es comparar el efecto de una condición sobre un grupo con el efecto de una condición diferente sobre un segundo grupo, o bien, comparar el efecto de condiciones diferentes sobre el mismo grupo. Para hacer estas comparaciones son necesarios al menos dos grupos o dos condiciones.

Tal vez, la manipulación directa de las variables independientes sea el rasgo más distintivo de la investigación experimental. *Manipulación*, en este sentido, significa que el investigador determina y controla el tratamiento o la condición específicos para cada grupo de sujetos. La variable independiente se manipula según esos diferentes valores o bien las condiciones (niveles) de la variable independiente son asignadas por el experimentador a los grupos. Si las condiciones no pueden ser asignadas por el investigador, el estudio no será entonces un experimento. Supongamos, por ejemplo, que un grupo de investigación está interesado en averiguar si el orden de dificultad de las cuestiones de un test afecta al rendimiento del estudiante. El estudio podría tener una variable independiente, el orden de las cuestiones, con dos niveles, ítems ordenados desde el más fácil hasta el más difícil e ítems ordenados desde el más difícil hasta el más fácil. Estas son las condiciones que manipulan los investigadores que, probablemente, dividirían una clase en dos grupos, de manera aleatoria y pasarían a los alumnos de cada grupo uno de los dos tipos de orden de las cuestiones. Es importante señalar que hay muchas variables en educación que no pueden manipularse nunca, como por ejemplo la edad, el peso, el sexo y la situación socioeconómica. Estas variables se denominan **variables asignadas** o **atribuidas** porque no pueden ser manipuladas. Aunque se pueden incluir variables asignadas en la investigación

experimental, debe haber, al menos, una variable manipulada para que la investigación sea considerada un experimento.

La cuarta característica de una investigación experimental, medición de las variables dependientes, significa que la investigación experimental está relacionada con aspectos a los que se puede asignar un valor numérico. Si el resultado del estudio no puede ser medido ni cuantificado de ningún modo, la investigación no podrá ser experimental.

Otra característica que implica números es el uso de la estadística inferencial. La estadística inferencial se utiliza para hacer enunciados de probabilidad sobre los resultados. Esto es importante por dos razones. Ya que en educación la medida es imperfecta y ya que, a menudo, queremos generalizar los resultados a grupos similares o a la población, la estadística inferencial nos permite realizar estas generalizaciones (ver el capítulo 9 en el que aparecen los detalles relacionados con la estadística inferencial).

La característica final de la investigación experimental es, quizá, la más importante desde un punto de vista general, porque el principio del control de las variables extrañas no es exclusivo de la investigación experimental. Lo que es característico en la investigación experimental es que hay un esfuerzo determinado para estar seguros de que ninguna variable extraña que pueda estar controlada proporciona hipótesis alternativas posibles para explicar los resultados. Controlamos las variables extrañas ya sea asegurándonos de que éstas no tienen ningún efecto en la variable dependiente o bien manteniendo el mismo efecto de las variables extrañas para todos los grupos. Este capítulo hace especial hincapié en esta característica de la investigación experimental. Además, muchos de los principios de control de las variables extrañas se aplican también a los diseños *ex post facto*.

Cualidades y limitaciones de la investigación experimental

El método experimental es, sin lugar a dudas, la mejor aproximación para determinar el efecto causal de una variable única y aislada sobre algo. Esto se explica, sobre todo, por el alto nivel de control de las variables extrañas y por el poder de manipulación de variables. Sin embargo, el control exhaustivo que caracteriza a una buena investigación experimental se convierte en una obligación en el campo educativo. El control se consigue más fácilmente en la investigación con personas sólo en medios restringidos y artificiales. Esto es una limitación en educación por dos motivos. Las personas reaccionan a las condiciones manipuladas y restringidas artificialmente de manera diferente al modo al que lo hacen ante condiciones que ocurren de manera natural; asimismo, si la investigación se realiza bajo condiciones artificiales, la generalizabilidad de los resultados (validez externa) está muy limitada.

He aquí un ejemplo de este dilema. El problema a investigar es si un enfoque individualizado o un grupo cooperativo de discusión es el mejor método para enseñar conceptos científicos a los alumnos de cuarto curso. El objetivo es encontrar el método de enseñanza que obtenga los mejores resultados de rendimiento. Se seleccionará un enfoque experimental porque el problema es un claro problema de causa y, presumiblemente, el método de instrucción —la variable independiente— puede manipularse fácilmente. Para maximizar el control de las variables extrañas se podría planear el experimento de la siguiente manera: al comienzo de una jornada escolar se conduce a todos los alumnos de cuarto curso a una habitación especial, donde se dividen de manera aleatoria entre grupos individualizados y equipos. Para eliminar cualquier efecto debido al profesor, actuarán licenciados en la universidad como profesores ayudantes. Para evitar cualquier repercusión de las diferentes habitaciones, se lleva a cada grupo de alumnos a habitaciones parecidas situadas en lugares diferentes. Para controlar las posibles distracciones, las habitaciones no tienen ventanas. Para asegurarse de que las instrucciones son uniformes, los profesores las leen de unos guiones preparados al efecto. El material de ciencias seleccionado ha sido cuidadosamente ocultado para que se trate de información nueva para los alumnos. Tras haber estudiado el material durante una hora, se examina al alumno sobre los conceptos. El formato de la prueba, el tiempo para completar las preguntas y otros procedimientos son los mismos para ambos grupos. Los resultados comparan el rendimiento de los grupos y, dado que el diseño ha controlado la mayoría de las variables extrañas, se podrá atribuir esta diferencia a la variable independiente, el método de instrucción.

¿Qué hacemos con estos conocimientos? Ya que un enfoque parecía mejor en este experimento, ¿significa esto que el señor Jones, en su clase, con su estilo de enseñanza, con alumnos que pueden tener aptitudes y carencias particulares, debería usar el supuesto método puesto a prueba? Quizá, pero la dificultad que se presenta es de generalizabilidad, un problema común en los experimentos que son capaces de demostrar un control exhaustivo de las variables extrañas, —es decir, la validez interna—. Por otra parte, si queremos maximizar la validez externa, necesitamos diseñar el experimento justo en la clase del señor Jones así como en las clases de otros profesores, y diseñar de algún modo el estudio para controlar tantas variables como nos sea posible, sin alterar el ambiente natural de la clase. El investigador necesitaría seleccionar las variables que pudieran, con más probabilidad, afectar al rendimiento, por ejemplo, aptitud, hora del día y composición de los grupos, y controlar éstas en la medida de lo posible. Este enfoque dificulta aún más la demostración de qué método de enseñanza es más eficaz, pero los resultados son más generalizables a las clases normales. El verdadero reto está en el diseño de las pruebas para que los resultados obtenidos puedan ser razonablemente generalizados a otras personas y ambientes, es decir, mantener un

equilibrio entre la validez interna y la externa en un diseño. Esta tarea es difícil pero no imposible, y uno de los objetivos de este capítulo es presentar varios diseños y procedimientos que permitan conclusiones razonables de causa-efecto en el contexto de escenarios naturales.

Esperamos que el empleo de la palabra *razonable* no resulte confuso. Simplemente, nos aproximamos al diseño experimental puro todo lo que podemos porque estos diseños determinan convincentemente la causalidad. Sin embargo, en el análisis final, ya que trabajamos con seres humanos en situaciones complejas, debemos casi siempre recurrir a la opinión de profesionales o a la razón para llegar a conclusiones a partir de los resultados obtenidos. La información relativa a los diseños reflejada en este capítulo y las amenazas a la validez interna y externa nos ayudan a realizar este juicio.

Finalmente, antes de presentar los diseños, sería importante señalar que la investigación experimental no es apropiada para todo tipo de investigación educativa; sólo lo es para algunas investigaciones que buscan información sobre las relaciones causa-efecto. El método experimental sería inapropiado para muchos problemas educativos como los estudios descriptivos, por ejemplo, «¿cuál es la actitud o el nivel de rendimiento?» o los estudios de las relaciones, por ejemplo, «¿existe una relación entre la edad y la autoestima?». En algunas situaciones, un enfoque cualitativo sería más válido para explicar los hechos, mientras que en estudios de evaluación se usan frecuentemente experimentos junto con otros enfoques para investigar cuestiones sobre una práctica determinada.

Planificación de una investigación experimental

Aunque la secuencia de pasos en una investigación experimental es básicamente la misma que en otros tipos de investigación, algunos procedimientos y características de la planificación son característicos del diseño experimental.

El primer paso es definir un tema de investigación, revisar la bibliografía y formular hipótesis de investigación claras. Es esencial que la investigación experimental se guíe por hipótesis de investigación que enuncien los resultados esperados. Los resultados reales respaldarán o no las hipótesis de investigación.

En el segundo paso, el investigador selecciona sujetos pertenecientes a una población determinada y, dependiendo del diseño específico utilizado, frecuentemente «asigna» sujetos a grupos diferentes. Un estudio experimental simple implica dos grupos, uno de ellos denominado **grupo experimental** o **grupo de tratamiento** y el otro considerado como **grupo de control** o **de comparación**. A continuación, a cada grupo se le asigna un nivel de la variable independiente. De manera técnica, un grupo de control no recibe ningún tratamiento (por ejemplo, cuando se comparan fumadores con no fumadores), pero en la mayoría de las investigaciones educativas no es productivo comparar un grupo que recibe un

tratamiento con otro grupo que no recibe nada. Sería como comparar niños que reciben una autorización especial e individualizada con niños que no la reciben para, a continuación, concluir basándose en los resultados que los niños necesitan una autorización individualizada. Tampoco es realista esperar que en los ambientes escolares haya un grupo que no haga nada mientras otro recibe un tratamiento especial. Por estas razones, es preferible concebir los dos grupos de la investigación experimental como grupos de tratamiento y de comparación, o como un grupo que recibe un método A y otro que recibe un método B, en vez de un grupo experimental y otro de control.

En la asignación de los tratamientos, indicándolos mediante niveles de la variable independiente, el investigador determinará la naturaleza del valor, las formas o las condiciones que cada grupo recibe. Esto puede ser una asignación simple como, por ejemplo, clase magistral frente a grupo de discusión, reprimendas en voz alta o en privado o estilo de liderazgo autocrítico frente a participativo; o bien, podría haber más de dos niveles, con grados variables de la condición en cada nivel de la variable independiente. Por ejemplo, si un investigador estuviera interesado en el efecto de los diferentes tipos de *feedback* o retroalimentación del profesor sobre las actitudes de los alumnos, entonces la retroalimentación, variable independiente, podría representarse según cuatro niveles: sólo la nota, la nota más una palabra, la nota más una frase y la nota más tres frases. Entonces, el investigador «constituiría» cuatro niveles mediante la asignación aleatoria de sujetos en cuatro grupos. El investigador llevaría a cabo un control apropiado para que ninguna diferencia en las actitudes pudiera considerarse causada por diferentes tipos de *feedback*, o para que ninguna diferencia en las actitudes indicara una relación no causal entre el tipo de retroalimentación y las actitudes.

Una de las dificultades, al planificar una investigación experimental, consiste en saber si los tratamientos serán lo suficientemente eficaces, es decir, si la condición del tratamiento es ofrecer retroalimentación a los alumnos, ¿será el impacto de la retroalimentación suficiente para afectar a las actitudes de los alumnos? ¿Supondría una diferencia proporcionar el *feedback* durante varios días? Quizás, es necesario proporcionar retroalimentación durante un mes o más. En otras palabras, o bien debe probarse el tratamiento con antelación para estar seguros de que tiene poder para provocar un impacto, o se debería dar tiempo suficiente al tratamiento para que funcionase. Este puede ser un problema especialmente difícil en muchas de las investigaciones educativas porque hay muchas influencias como el rendimiento, las actitudes, la motivación y la autoestima que pueden afectar a las variables dependientes, y resulta difícil determinar una variable independiente específica que tenga un efecto significativo y único dadas todas las demás influencias. Finalmente, los tratamientos experimentales, a veces, no son lo bastante diferentes de los realizados a los grupos de

comparación para que sea posible una diferencia estadística. Por ejemplo, en un estudio sobre diversas técnicas de orientación, si la única diferencia entre las condiciones experimentales y de control es la distancia a la que se sienta el orientador de sus clientes (es decir, cuatro o seis pasos) es improbable, dadas todas las otras influencias, que un investigador obtenga una diferencia en los resultados.

Una vez que se han establecido las condiciones del tratamiento, es necesario especificar el diseño que se empleará. Este capítulo presenta la mayoría de los diseños básicos utilizados en la investigación educativa experimental. Los diseños incluyen los procedimientos para la asignación de sujetos, el número de grupos y el momento en que se administrará el tratamiento a cada grupo. La principal preocupación del investigador a la hora de elegir un diseño es maximizar la validez interna, un concepto al que se volverá en la próxima sección.

VALIDEZ INTERNA DE LOS EXPERIMENTOS

En el capítulo 4 comentábamos el concepto de validez interna. La validez interna de un estudio es un juicio que se hace con respecto a la confianza con la que las hipótesis alternativas posibles pueden ser descartadas como explicaciones en los resultados. Esto implica un proceso deductivo en el que los investigadores deben examinar de manera sistemática cómo, cada una de las amenazas a la validez interna que constituyen las hipótesis alternativas rivales, pueden haber influido en los resultados. Si todas las amenazas pueden ser eliminadas de manera razonable, entonces el investigador puede estar seguro de que una relación observada es causal y de que la diferencia en las condiciones del tratamiento causaron los resultados obtenidos. La validez interna es raramente una decisión de todo o nada. Más bien se la evalúa como una cuestión de grado, dependiendo de la posibilidad de la explicación. Como se precisará más adelante, algunos diseños, son relativamente firmes con respecto a la validez interna, porque la mayoría de las hipótesis alternativas pueden ser descartadas sin problema, mientras otros diseños que se prestan a un conjunto de explicaciones rivales verosímiles son débiles con respecto a la validez interna. En el análisis final, los investigadores deben ser sus mejores críticos y examinar detenidamente todas las amenazas imaginables. Por lo tanto, es importante para los usuarios de la investigación así como para aquéllos que la realizan ser conscientes de las amenazas comunes a la validez interna y de los mejores modos de controlarlas.

Una manera útil de pensar en las posibles amenazas a la validez interna es distinguir lo que es «posible» de lo que es «verosímil». Una posible amenaza está claramente fuera del control del diseño, pero para que sea considerada en serio, la posible amenaza tiene que ser verosímil en dos sentidos. En primer lugar, tiene que ser algo que afecte potencialmente a la variable dependiente. Por ejemplo, puede haber diferencia entre los grupos comparados por el color de sus ojos,

¡pero no es probable que este hecho se encuentre relacionado con la variable dependiente! En segundo lugar, tiene que estar sistemáticamente relacionado con un grupo. Es decir, tiene que afectar a un grupo más que al otro. Si un factor afecta a ambos grupos por igual, entonces, en general, no supone una amenaza a la validez interna.

Las amenazas a la validez interna presentadas en el capítulo 4 están resumidas en la tabla 8.1 para ayudar a recordar cada factor. Éstos se discutirán de nuevo en el contexto de cada diseño resumido en las próximas cuatro secciones del

TABLA 8.1: RESUMEN DE AMENAZAS A LA VALIDEZ INTERNA

Amenaza	Descripción
Historia	Sucesos no planificados o extraños que ocurren durante la investigación y afectan a los resultados.
Selección	Las diferencias entre los sujetos en los grupos pueden dar lugar a resultados diferentes debido a la composición de los grupos.
Regresión estadística	Puntuaciones de grupos de sujetos que se aproximan a la media debido a que los sujetos han sido identificados por sus puntuaciones extremadamente altas o bajas.
Efecto del pretest	Ocurre cuando el acto de realizar una prueba o responder a un cuestionario anterior al tratamiento afecta a los sujetos.
Instrumentación	Diferencias en los resultados debido a la falta de fiabilidad, cambios en las pruebas de medida o en los observadores.
Mortalidad experimental	Ocurre a causa de una pérdida sistemática de sujetos.
Maduración	Ocurre cuando hay un efecto debido a la maduración o a otros cambios naturales en los sujetos (por ejemplo, ser más mayor, más listo, más fuerte, estar más cansado).
Difusión del tratamiento	Ocurre cuando los sujetos de un grupo aprenden de tratamientos o condiciones de grupos diferentes.
Efectos del experimentador	Efectos deliberados o involuntarios del investigador en las respuestas del sujeto.
Repeticiones del tratamiento	El número de repeticiones del tratamiento es diferente al número de sujetos.
Efectos del sujeto	Cambios en el comportamiento como respuesta a ser un sujeto experimental o a estar en un experimento.
Conclusión estadística	Violación de las presunciones o uso incorrecto de las pruebas estadísticas.

capítulo, tanto en el texto como en las tablas desde la 8.3 hasta la 8.5, que ofrecen un repaso de las amenazas que cada diseño controla.

VALIDEZ EXTERNA DE LOS EXPERIMENTOS

La validez externa es el grado en que los resultados de un experimento pueden generalizarse a personas y condiciones ambientales fuera del contexto del experimento. Es decir, si el mismo tratamiento o condiciones experimentales se reprodujeran con sujetos diferentes en un escenario diferente, ¿serían iguales los resultados? En otras palabras, ¿cuáles son las características de los sujetos y de las condiciones ambientales para las que podemos esperar los mismos resultados? Concluimos que un experimento tiene una validez externa firme si la generalizabilidad es relativamente grande, mientras que tiene una validez externa débil si no somos capaces de generalizar mucho más allá del experimento real. La tabla 8.2 resume las fuentes de las amenazas a la validez externa (ver la revisión de las amenazas a la validez externa en el capítulo 4).

Es difícil considerar la validez externa del mismo modo que observamos la validez interna porque la mayoría de los experimentos no están concebidos específicamente para controlar las amenazas a la validez externa. Los investigadores controlan conscientemente algunas amenazas a la validez interna utilizando un diseño particular, pero la mayoría de las amenazas a la validez externa no están relacionadas con el diseño. En sólo algunos diseños se puede llegar a la conclusión de que las fuentes de la validez externa están controladas. Por ejemplo, bajo la categoría de validez ecológica, amenazas tales como la descripción de variables, el efecto de novedad, la interacción entre la situación y el tratamiento o la interacción entre el tiempo de medición y el tratamiento no están controladas por ningún diseño experimental en particular. Estas amenazas están más en función de los procedimientos y definiciones que de un diseño y, en la mayoría de los estudios, el lector decide si alguna de las amenazas es razonable.

Dos de las amenazas ecológicas, la interferencia de múltiples tratamientos y la sensibilización al pretest y al postest están sólo presentes en diseños concretos. La interferencia de tratamientos múltiples sólo se considera si se aplica más de un tratamiento de manera secuencial. La sensibilización al pretest se convierte en una amenaza seria cuando se investigan personalidad, valores, actitudes u opiniones, ya que realizar un pretest puede sensibilizar al sujeto con respecto al tratamiento. Por ejemplo, si un estudio que investiga el efecto de un taller sobre las actitudes acerca del uso de ordenadores en el ámbito educativo hiciera un pretest de actitudes, el propio pretest podría afectar a las consiguientes actitudes independientemente del taller o en interacción con el mismo.

La interacción pretest-tratamiento se minimiza en estudios realizados durante mucho tiempo, como varias semanas o meses, así como en estudios con niños

TABLA 8.2: RESUMEN DE AMENAZAS A LA VALIDEZ EXTERNA

Amenaza	Descripción
Población Selección de sujetos Características de los sujetos Interacción sujeto-tratamiento	La generalización se limita a los sujetos de la muestra si éstos no han sido seleccionados de manera aleatoria a partir de una población identificada. La generalización se limita a las características de la muestra o de la población (por ejemplo, estatus socioeconómico, capacidad, edad, lugar, raza). La generalización puede estar limitada a causa de la interacción entre los sujetos y el tratamiento (es decir, el efecto del tratamiento sólo afecta a algunos sujetos).
Ecológica Descripción de las variables Interferencia de tratamientos múltiples Interacción situación-tratamiento Interacción tiempo de medición-tratamiento Sensibilización pretest y postest Efecto de novedad o de disrupción	La generalización se limita a las definiciones operacionales de las variables dependientes e independientes. En experimentos en los que los sujetos reciben más de un tratamiento, la generalizabilidad se limita a las situaciones parecidas de tratamiento múltiple, a causa del efecto del primer tratamiento en los tratamientos sucesivos. La generalización se limita a la situación en la que el estudio se realiza (por ejemplo, la habitación, el momento del día, otras personas presentes, otros lugares cercanos). Los resultados pueden limitarse al periodo de tiempo en el que fueron obtenidos. Los tratamientos que causan efectos inmediatos podrían carecer de efectos duraderos. El pretest o el postest pueden interactuar con el tratamiento de forma que se obtendrían resultados parecidos sólo cuando las circunstancias de la prueba estuvieran presentes. Los sujetos pueden responder de manera diferente debido a un cambio en su rutina y la generalización puede limitarse a situaciones que implican novedades o interrupciones parecidas (por ejemplo, un tratamiento en principio eficaz podría llegar a ser inefectivo una vez que la novedad pasara).

pequeños. La sensibilización al postest ocurre sólo si éste sensibiliza al sujeto sobre el tratamiento y afecta a los resultados de modo que no habrían ocurrido sin el postest.

A menudo, los investigadores educativos se enfrentan al difícil dilema de que si se maximiza la validez interna, hay que sacrificar la validez externa. Una elevada validez interna requiere un control estricto de todas las fuentes de las va-

riables de confusión, un tipo de control que podría llevar al estudio a condiciones de laboratorio. Sin embargo, cuanto más se controla el entorno, los resultados son menos generalizables a otras situaciones. Este es un dilema constante para los educadores porque la investigación que no puede ser utilizada para otras poblaciones en otras situaciones contribuye muy poco a la práctica educativa, pero debe haber un control suficiente como para extraer conclusiones causales razonables. Y, por supuesto, sin la validez interna, la validez externa es una cuestión discutible. La mayoría de los investigadores se esfuerzan por equilibrar las amenazas a la validez interna y externa utilizando el rigor suficiente para conseguir unos resultados que estén justificados científicamente y realizando el estudio bajo condiciones que faciliten la generalización a otras situaciones. Un buen enfoque para resolver el dilema es repetir estudios exhaustivamente controlados con poblaciones diferentes en situaciones diferentes.

DISEÑOS PREEXPERIMENTALES

Los tres diseños que presentaremos en esta sección se denominan **diseños preexperimentales** porque no cumplen con dos o más de las seis características de la investigación experimental enumeradas anteriormente. Como consecuencia, pocas amenazas a la validez interna están bajo control. Esto no significa que estos diseños no sean interpretables, ni tampoco que no deban utilizarse. Hay algunos casos en los que pueden descartarse amenazas sobre la base de una teoría aceptada, el sentido común u otros datos. Sin embargo, como no consiguen descartar la mayoría de las hipótesis alternativas, resulta difícil realizar deducciones causales razonables a partir, exclusivamente, de estos diseños. Tal vez, éstos se usen mejor como un modo de generar ideas que puedan ser puestas a prueba de manera más sistemática.

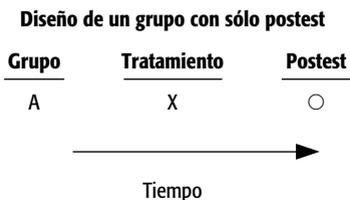
Notación

Al presentar los diseños en este capítulo usaremos un sistema de notación que ofrece información que facilita el entendimiento de los diseños. Este sistema de notación es propio, aunque es parecido al sistema básico utilizado por Campbell y Stanley (1963) y Cook y Campbell (1979). Nuestro sistema es el siguiente:

- R Asignación aleatoria.
- O Observación, una medida que registra observaciones de un pretest o de un postest.
- X Condiciones de tratamiento (el subíndice 1 hasta n indica diferentes tratamientos).
- A, B, C, D, E, F Grupos o sujetos o, para los diseños caso único, línea base o condiciones de tratamiento.

Diseño de un grupo sólo con posttest

En el **diseño de un grupo sólo con posttest** el investigador aplica un tratamiento y, a continuación, hace una observación, como se representa en el diagrama siguiente, donde A es el grupo de tratamiento, X es el tratamiento y O es el posttest.



Dado que no todas las amenazas a la validez interna son aplicables a este diseño, porque no hay ningún pretest y ninguna comparación con otros tratamientos, sólo se pueden hacer como máximo tentativas de conclusiones causales. Por ejemplo, sin un pretest, es difícil llegar a la conclusión de que un comportamiento ha cambiado por completo (por ejemplo, cuando se prueba un método de enseñanza de matemáticas a los alumnos que conocen las respuestas del examen final antes de recibir ningún tipo de enseñanza). Sin una comparación o grupo de control es difícil también saber si otros factores que ocurren al mismo tiempo que el tratamiento estaban relacionados de manera causal con la variable dependiente. Más incluso, aunque sólo cinco de las amenazas a la validez interna resultan relevantes en este diseño, las limitaciones antes reflejadas son tan graves que los resultados de la investigación basados en este diseño por sí solos son, con frecuencia, imposibles de interpretar (ver la tabla 8.3). La única situación en la que este diseño es razonable es cuando el investigador puede estar seguro del nivel de conocimiento, actitud o destreza de los sujetos antes del tratamiento y puede estar convencido de que la historia no es una amenaza. Por ejemplo, veamos el caso de un profesor de métodos de investigación que quiere realizar un estudio sobre lo que los alumnos han aprendido sobre regresión estadística en sus clases. Parece razonable llegar a la conclusión de que éstos no sabían demasiado sobre regresión antes de que comenzara el curso y de que es improbable que aprendieran por otros medios, por ejemplo, ¡durante reuniones con amigos! Por consiguiente, el diseño de un grupo sólo con posttest puede ofrecer resultados válidos.

No debería confundirse este diseño con el llamado «diseño de estudio de un caso». Los diseños de estudios de caso, tal como los hemos definido, son investigaciones cualitativas sobre una persona, grupo, acontecimiento o situación en un periodo determinado. Estos estudios realizan exámenes detallados de fenómenos complejos (para más información, ver capítulo 10).

Diseño de un grupo con pretest-postest

Este diseño común se distingue del diseño de grupo sólo con postest por una única diferencia –la introducción de una observación que ocurre antes de se experimente la condición de tratamiento (pretest)–.

En el **diseño de un grupo con pretest-postest** se realiza a un grupo un pretest (○), a continuación el tratamiento (X) y después el postest (○). El pretest y el postest son los mismos, pero realizados en momentos diferentes. El resultado que se analiza es un cambio del pretest al postest. (A este diseño se le denomina popularmente diseño pretest-postest). Aunque el investigador puede obtener al menos una medida de cambio con este diseño hay muchas hipótesis alternativas verosímiles que son aplicables.

Diseño de un grupo con pretest-postest

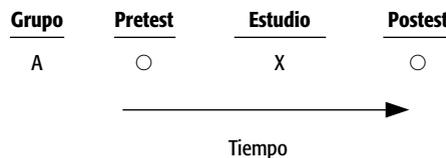


TABLA 8.3: AMENAZAS A LA VALIDEZ INTERNA DE DISEÑOS PREEXPIMENTALES

Diseño	Historia	Selección	Regresión estadística	Pretest	Instrumentación	Mortalidad experimental	Maduración	Difusión del tratamiento	Efectos del experimentador	Repeticiones del tratamiento	Efectos del sujeto	Conclusión estadística
1. Un grupo sólo con postest	–	–	?	SA	?	?	–	SA	?	?	?	SA
2. Un grupo con pretest-postest	–	?	?	–	?	?	–	SA	?	?	?	?
3. Grupos no equivalentes sólo con postest	?	–	?	SA	?	?	?	?	?	?	?	?

Nota: En esta tabla, al igual que en las tablas 8.4 y 8.5, un signo menos significa una limitación; un signo más que el factor está controlado; un signo de interrogación una posible fuente de invalidez y SA significa que la amenaza no es aplicable a este diseño.

Consideremos este ejemplo. Un profesor de universidad ha recibido una ayuda económica para realizar talleres dirigidos a maestros sobre el tema de la integración escolar. Uno de los objetivos del programa es la mejora de las actitudes de los maestros hacia la integración de los niños con discapacidad. Para evaluar este objetivo, el profesor selecciona un diseño pretest-postest, administrando a los maestros, antes de los talleres, una encuesta sobre actitudes a modo de pretest y volviendo a pasar la misma encuesta de nuevo tras la celebración de los talleres (postest). Suponga que las puntuaciones del postest son más altas que las del pretest, ¿puede concluir el investigador que la causa en el cambio de las puntuaciones ha sido el taller? Tal vez, pero existen varias amenazas posibles a la validez interna y, hasta que no sean eliminadas, el investigador no puede asumir que la asistencia al taller fue la causa del cambio.

La amenaza más seria es la relativa a la historia. Al no existir un grupo de control o de comparación, el investigador no puede estar seguro de que otros sucesos que hayan ocurrido entre el pretest y el postest no hayan causado el cambio actitudinal. Algunos de estos sucesos han podido suceder en el contexto del taller (un maestro cita un testimonio acerca de niños excepcionales) y otros pueden ocurrir fuera del contexto del taller (durante el taller aparece en el periódico escolar un artículo sobre integración). Hechos como éstos no están controlados y pueden afectar a los resultados. Es necesario, pues, que el investigador elabore un caso que justifique que esos efectos no son probables o que si lo son, no afectan. A veces se emplean los datos para descartar la existencia de determinadas amenazas, pero en muchos casos, se emplea sencillamente el sentido común, la teoría o la experiencia. Y aunque se presente, si una posible amenaza no ha podido ser eliminada, el investigador debe admitir que el efecto del taller en el cambio de actitudes no puede ser determinado con exactitud.

La selección no es una amenaza a la validez interna en este diseño porque sólo se estudia un grupo, pero las características de los sujetos podrían interactuar con el tratamiento y afectar a la validez externa de los resultados.

La regresión estadística podría ser un problema en este diseño si los sujetos han sido seleccionados con el criterio de puntuaciones extremadamente altas o bajas. Por ejemplo, en nuestro supuesto sobre el taller, imagine que el director del colegio quisiera sólo a aquellos maestros con actitudes menos favorables. Las puntuaciones del pretest serían en este caso muy bajas y, a causa de la regresión, podrían ser más altas en el postest con independencia del efecto del taller.

El efecto de la realización de un pretest es, a menudo, una amenaza para la investigación realizada con este diseño, especialmente para la investigación sobre actitudes, ya que el simple hecho de realizar un pretest puede alterarlas. El contenido del cuestionario puede sensibilizar a los sujetos hacia problemas específicos o puede aumentar el nivel de conciencia general del asunto en cuestión y provocar que reflexionen más acerca del tema. La instrumentación también

puede constituir una amenaza. Si, por ejemplo, los maestros realizan un pretest el viernes por la tarde y el postest el miércoles siguiente por la mañana, las respuestas pueden ser diferentes simplemente a causa de las actitudes generales, que es probable que prevalezcan en cada uno de esos momentos del día y de la semana.

La mortalidad experimental puede ser un problema si, durante el pretest y el postest, se pierden sujetos por alguna razón. Si todos los maestros de una escuela comienzan un taller, por ejemplo, y aquéllos que tienen actitudes más negativas hacia la integración abandonan porque no quieren aprender más acerca del tema, las medidas de la actitud de los demás sujetos serán, en este caso, altas. Considere otro ejemplo. Para evaluar el efecto de las iniciativas de una escuela para reforzar las actitudes favorables hacia el aprendizaje, los alumnos hacen un pretest en su segundo año y un postest en el último. Un posible argumento, al menos uno que debería ser tenido bajo control, es que la mejora en las actitudes queda demostrada porque los alumnos que poseían, en principio, las actitudes más negativas nunca llegaron a finalizar sus estudios y abandonaron. La mortalidad experimental es especialmente problemática en los casos en los que se emplean poblaciones «en tránsito», en los experimentos de larga duración y en la investigación longitudinal.

La madurez es una amenaza a la validez interna de la investigación basada en este diseño cuando la variable dependiente es inestable a causa de cambios en la madurez. Esta amenaza es más seria conforme se incrementa el periodo entre el pretest y el postest. Suponga, por ejemplo, que un investigador está estudiando la autoestima de los alumnos de los últimos cursos de primaria. Si el tiempo que transcurre entre el pretest y el postest es relativamente corto (dos o tres semanas), probablemente la madurez no será un problema; pero si transcurre un año entre el pretest y el postest, se producirán con toda probabilidad cambios en la autoestima independientemente del tratamiento y debidos a la madurez. La maduración incluye amenazas tales como estar paulatinamente más cansado, aburrido o hambriento durante la administración de las pruebas, lo que puede llegar a constituir un problema en algunos diseños pretest-postest. En el ejemplo de los talleres de integración, es improbable que la maduración sea una amenaza seria y, probablemente, resulta razonable descartarla como hipótesis rival verosímil.

Dado que en este diseño existe sólo un tratamiento, la difusión no es aplicable. La repetición de los tratamientos puede llegar a ser una amenaza, dependiendo del modo en que sean administrados. Las amenazas de los efectos del experimentador, los del sujeto y de la conclusión estadística son posibles en cualquier experimento y deberían ser tenidos en cuenta.

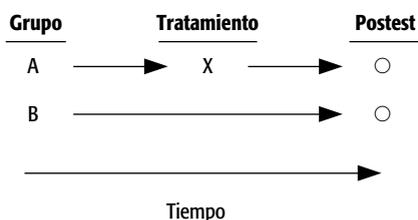
De lo anteriormente descrito se deduce que existen muchas amenazas a la validez interna, no controladas, en un diseño de un grupo con pretest-postest. En consecuencia, este diseño sólo debería ser empleado en contadas condiciones que

minimicen la posibilidad de existencia de las amenazas (por ejemplo, usando pruebas fiables y empleando cortos intervalos de tiempo entre el pretest y el posttest) y cuando resulta imposible emplear otros diseños que puedan controlar algunas de estas amenazas.

Grupos no equivalentes sólo con posttest

Este diseño es parecido al de un grupo sólo con posttest. La diferencia es que en un **diseño de grupos no equivalentes sólo con posttest** hay un grupo de control o de comparación que no recibe tratamiento alguno o un tratamiento diferente se añade al diseño de un grupo sólo con posttest. El diseño se refleja en el siguiente diagrama.

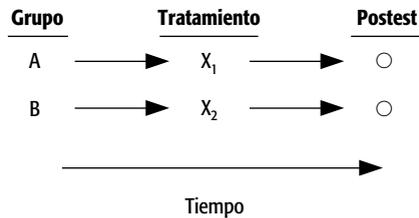
Diseño de grupos no equivalentes sólo con posttest



Este diseño se utiliza a menudo tras implementar un tratamiento. El procedimiento consiste en aplicar el tratamiento a un grupo y, a continuación, evaluar la variable dependiente (mediante el posttest), y realizar sólo el posttest a otro grupo al mismo tiempo que se realiza el posttest al primer grupo. El término *grupos no equivalentes* se utiliza como nombre para el diseño porque la selección es la amenaza más seria a la validez interna de los resultados. Advierta que no se realiza una asignación aleatoria de sujetos a los grupos. Por lo tanto, las diferencias entre los grupos de los sujetos puede tener peso en las diferencias en los resultados de los posttests. Cuanto más diferentes sean los grupos, es más verosímil que la selección se convierta en la causa de los resultados. Supongamos, por ejemplo, que el profesor que dirigía el taller sobre integración quería conseguir un grupo de comparación y localizó una escuela que quería colaborar. Incluso, aunque los resultados posttest del grupo de tratamiento fueran mejores que los del grupo de comparación, no se podría concluir que los mejores resultados se debieron al taller. Podría ser que los profesores en la escuela de tratamiento tuvieran una actitud más favorable al empezar el estudio y que el taller influyera poco sobre las actitudes de sus profesores.

Existen también otras amenazas menos serias a la validez interna de la investigación basadas en este diseño. Estas amenazas aparecen cuando el diseño básico incluye estudios alternativos, como se indica a continuación.

**Diseño de grupos no equivalentes con
tratamiento alterno sólo con postest**



Este diseño se utiliza cuando un investigador quiere comparar dos o más tratamientos pero no puede realizar un pretest o aleatorizar la asignación de sujetos a cada grupo. En este caso interno, la historia interna o intragrupo es una amenaza, ya que lo que podría suceder dentro de cada grupo, sin estar relacionado con el tratamiento, podría afectar al postest. Normalmente, la historia externa no es una amenaza, a menos que diferencias en la selección expongan a los sujetos fuera del contexto del tratamiento a diferentes condiciones que afecten a los resultados. La regresión puede ser una amenaza aunque sólo se haga una observación. El efecto del pretest no es una amenaza porque no hay pretest, pero la instrumentación puede ser una amenaza si existen diferencias en la manera en la que se evalúa el postest en cada grupo (como, por ejemplo, si el observador de un grupo está más atento que el del otro). La mortalidad experimental es una amenaza a causa de la pérdida de sujetos, debido a que bien las características iniciales de los sujetos (selección) como diferentes tratamientos podrían hacer que algunos sujetos abandonasen. La maduración puede ser también una amenaza, dependiendo de las características de la selección. Si los sujetos de cada grupo conocen el tratamiento administrado al otro grupo, es posible que la difusión del estudio represente una amenaza. También pueden incidir los efectos del experimentador, los efectos del sujeto, las repeticiones del tratamiento y la conclusión estadística.

El diseño de grupos no equivalentes sólo con postest resulta relativamente débil para la comprobación de la causa. Si se utiliza este diseño, el investigador debería hacer grandes esfuerzos para conseguir grupos comparables y disminuir así la amenaza de la selección.

Las posibles fuentes de invalidez para la investigación realizada mediante estos tres diseños preexperimentales aparecen resumidas en la tabla 8.3. Como los diferentes diseños controlan diferentes factores y también tienen deficiencias particulares, el investigador elige el mejor diseño sobre la base de las condiciones de investigación. Si, por ejemplo, el investigador puede argumentar razonablemente que dos grupos se parecen con respecto a las variables importantes (nivel socioeconómico, rendimiento, edad, experiencia, motivación etc.), entonces, el mejor diseño sería el de grupos no equivalentes sólo con postest. En cualquier caso, todos estos diseños son relativamente débiles para ser usados con el

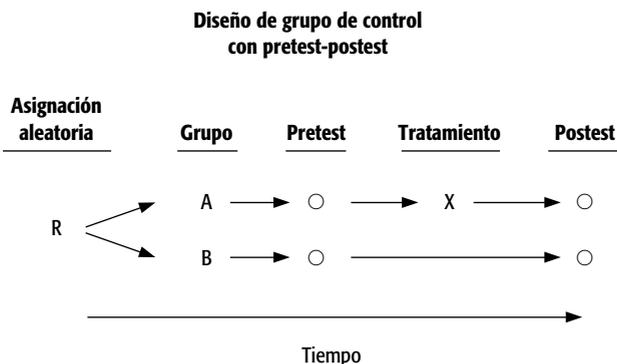
objeto de comprobar relaciones causales y, con frecuencia, con un poco de planificación pueden modificarse ligeramente para permitir la realización de deducciones causales más razonables.

DISEÑOS EXPERIMENTALES¹

Esta sección presenta dos diseños que reciben el nombre de **diseños experimentales**. Ambos incluyen procedimientos para descartar diferencias entre los sujetos por medio de su asignación aleatoria a los grupos y ambos incluyen la manipulación de la variable de tratamiento. Estos diseños representan lo que históricamente ha sido denominado experimental en la biología y la física.

Diseño de grupo de control con pretest-posttest

Este **diseño de grupo de control con pretest-posttest** es una continuación del diseño de un grupo con pretest-posttest de dos modos: se añade un segundo grupo, denominado grupo de control o de comparación; y los sujetos son asignados a cada grupo de forma aleatoria. Este diseño se representa a continuación. El grupo A es el grupo experimental.



El primer paso es la asignación aleatoria de los sujetos al grupo experimental y al grupo de control. Normalmente, en estudios con un número relativamente pequeño de sujetos es mejor ordenarlos según el rendimiento, las actitudes u otros factores que puedan estar relacionados con la variable dependiente. Entonces, en el caso de un diseño de dos grupos, se forman parejas de sujetos; el investigador asigna de manera aleatoria un sujeto de cada pareja al grupo experimental y el otro sujeto al grupo de control. Otro procedimiento es emparejar

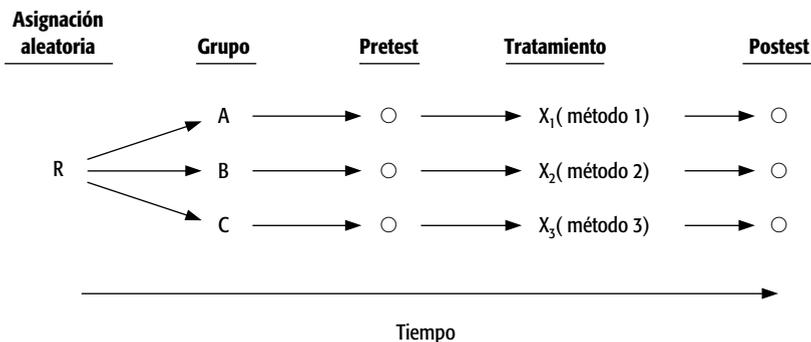
¹ Un tercer diseño experimental que aparece a menudo es el diseño de cuatro grupos de Solomon.

sujetos sobre la base de una variable y, a continuación, asignar miembros de cada grupo elegido a los grupos experimental y de control. Pero el propósito de la asignación aleatoria es permitir al investigador descartar de forma razonable todas las diferencias encontradas entre los grupos que pudieran pasar como diferencias encontradas en los resultados. Por lo tanto, con un pequeño grupo de sujetos es menos probable que los grupos sean iguales. Por ejemplo, si se asignan al azar tan sólo diez sujetos a dos grupos, existen muchas probabilidades de que, aunque la asignación sea aleatoria, haya importantes diferencias entre los grupos. Sin embargo, si doscientos sujetos son asignados aleatoriamente a los grupos, hay muy pocas probabilidades de que éstos difieran entre sí. En general, a los investigadores educativos les gusta tener al menos quince sujetos en cada grupo para asumir equivalencia estadística y confían más en los resultados si tienen de veinte a treinta sujetos en cada grupo.

El segundo paso es someter a un pretest a cada grupo sobre la variable dependiente (en algunos diseños se realiza primero el pretest, seguida de la asignación aleatoria). El tercer paso es administrar la condición de tratamiento al grupo experimental, pero no al grupo de control, manteniendo iguales el resto de las condiciones para ambos grupos, de modo que la única diferencia sea la manipulación de la variable independiente. A continuación, a cada grupo se le administra el postest de la variable dependiente.

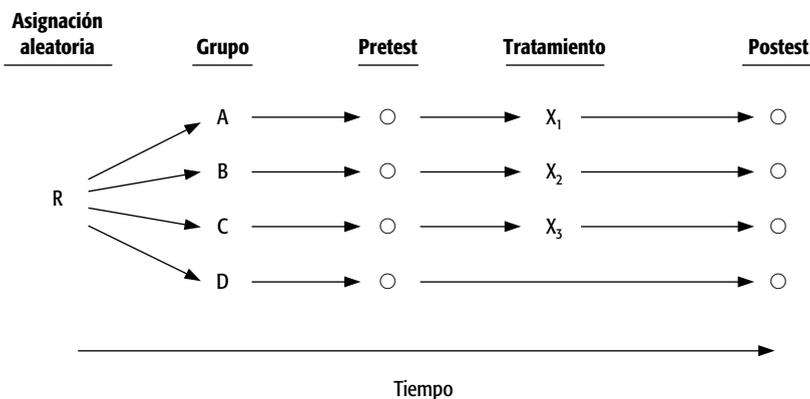
En el diseño del grupo de control reflejado en el diagrama, no hay tratamiento alguno destinado al grupo de control. Como indicamos anteriormente, es más común y, a menudo, deseable, contar con grupos de comparación en vez de grupos de control. Un diseño de comparación utiliza dos o más variaciones de la variable independiente y puede emplear dos o más grupos. Por ejemplo, supongamos que un profesor quiere comparar tres métodos de enseñanza de la ortografía. El profesor asigna aleatoriamente cada alumno de la clase a uno de los tres grupos, realiza un pretest, prueba los diferentes métodos y realiza un postest. El diseño tendría la siguiente apariencia:

Diseño de grupo de comparación con pretest-postest



Sería también posible combinar varios tratamientos diferentes con un grupo de control, como se muestra a continuación.

Diseño de grupo de control/comparación con pretest-postest



El diseño de grupo de control con pretest-postest controla cuatro fuentes de amenazas a la validez interna, como se indica en la tabla 8.4. Las amenazas relacionadas con la historia están, en general, controladas como hechos externos al estudio que afectan por igual a todos los grupos. Sin embargo, la razón para el signo de interrogación en la categoría correspondiente es que siempre es posible que puedan darse hechos únicos dentro de cada grupo y afectar a los resultados. La selección y la maduración están controladas a causa de la asignación al azar de sujetos. La regresión estadística y el efecto del pretest están controlados ya que cualquier efecto de estos factores es igual para todos los grupos. La instrumentación no es un problema cuando se utilizan los mismos procedimientos estandarizados, pero los estudios que utilizan observadores o codificadores deben intentar evitar los sesgos provenientes de éstos (saber qué alumnos están recibiendo un tratamiento determinado, emplear diferentes observadores o codificadores para cada grupo). Normalmente, la mortalidad experimental no es una amenaza a menos que un estudio particular provoque el abandono sistemático de sujetos.

La difusión de los tratamientos puede ser una fuente de invalidez en los experimentos en los que los sujetos de un grupo, a causa de una estrecha proximidad física o comunicación con los sujetos de otro grupo, conocen información de un tratamiento no destinado a ellos. Como en ese caso las condiciones destinadas a un grupo se transmiten a otros, se dispersa el efecto del tratamiento y no puede ser correctamente evaluado. Por ejemplo, si un investigador compara dos métodos de instrucción –instrucción de grupo cooperativo e instrucción individualizada– y realiza el experimento dentro de una clase de cuarto asignando aleatoriamente la mitad de la clase a cada método, es probable que los alumnos de un grupo sepan lo que está ocurriendo en el otro. Si los alumnos del grupo individualizado

TABLA 8.4: AMENAZAS A LA VALIDEZ INTERNA DE LOS DISEÑOS EXPERIMENTALES

Diseño	Historia	Selección	Regresión estadística	Pretest	Instrumentación	Mortalidad experimental	Maduración	Difusión del tratamiento	Efectos del experimentador	Repeticiones del tratamiento	Efectos del sujeto	Conclusión estadística
Grupo de control con pretest-postest	?	+	+	+	?	?	+	?	?	?	?	?
Grupo de control sólo con postest	?	+	SA	SA	?	?	+	?	?	?	?	?

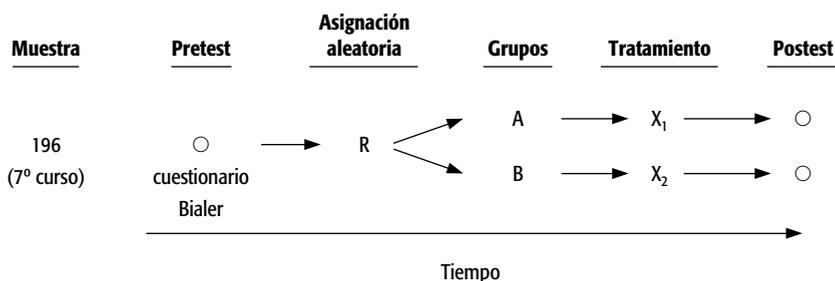
Nota: En esta tabla, al igual que en las tablas 8.3 y 8.5, un signo menos significa una limitación; un signo más que el factor está controlado; un signo de interrogación, una posible fuente de invalidez y SA significa que la amenaza no es aplicable a este diseño.

se sienten excluidos o creen que tienen una tarea menos interesante, podrían sentirse molestos y no dar tan buenos resultados. La divulgación podría también ocurrir si los alumnos del grupo cooperativo aprendieran a ayudar a los otros y, a continuación, apoyaran a los del grupo individualizado.

El efecto del experimentador constituye otra amenaza que podría suponer un problema, dependiendo del procedimiento del estudio. Si los individuos responsables de implementar los tratamientos son conscientes del propósito y de las hipótesis del estudio, podrían actuar de manera diferente con cada grupo y afectar así a los resultados. Si un profesor está implicado en un estudio que investiga el efecto que pueden tener cantidades diferentes de elogios en el comportamiento (a más elogios, mejor comportamiento), y sabe cuál debería ser el resultado hipotético, entonces, el profesor podría mostrar una actitud más positiva hacia el estudiante que recibe más elogios (mayor proximidad física, mayor contacto visual, actitud menos crítica) y así, contaminar el efecto esperado por la cantidad de elogios.

De forma similar, los efectos del sujeto pueden ser importantes si los sujetos de grupos diferentes respondieran de manera distinta a causa de sus respectivos tratamientos, por ejemplo, los sujetos que saben que están en el grupo de control podrían esforzarse o, por el contrario, podrían estar desmotivados porque no fueron seleccionados para el tratamiento «especial»; también puede ocurrir que los del tratamiento «especial» se sientan obligados a esforzarse más para dar «mejores» respuestas. Las repeticiones del tratamiento pueden ser una amenaza, dependiendo de cómo se administren éstos y la conclusión estadística es siempre una posibilidad.

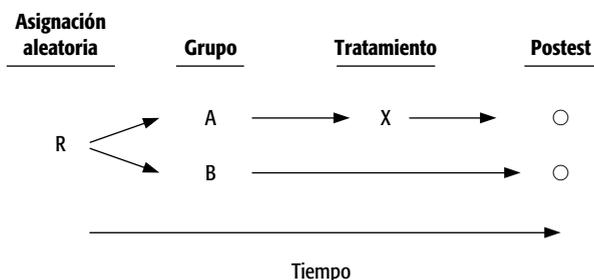
El extracto 8.1 ilustra un diseño de grupo de control con pretest-postest. En este estudio los investigadores administraron la misma prueba como pretest y postest. Para asegurar la misma representación en cada grupo de los sujetos con diferentes puntuaciones en el pretest, los resultados del pretest estaban ordenados por niveles de puntuación y los sujetos fueron asignados aleatoriamente de cada nivel a los grupos experimental o de control. Por lo tanto, se modificaron los pasos en la investigación ya que la asignación aleatoria tuvo lugar tras el pretest, y no antes. El diseño utilizado en el extracto 8.1 se representa a continuación.



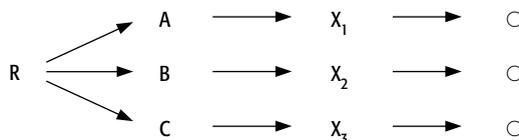
Diseño de grupo de control sólo con postest

El propósito de la asignación aleatoria, como se precisó anteriormente, es igualar los grupos experimental y de control antes de introducir la variable independiente. Si se equiparan los grupos a través de la asignación aleatoria, ¿es necesario realizar un pretest? Mientras hay casos en los que es preferible usar un pretest con asignación aleatoria, si los grupos tienen al menos quince sujetos cada uno, el pretest podría no ser necesario, es decir, el pretest no es esencial para realizar un estudio experimental auténtico. El **diseño de grupo de control sólo postest** es exactamente el mismo que el diseño de grupo de control del pretest y postest con la excepción de que no hay pretest en la variable dependiente. El diseño de grupo de control sólo con postest se representa de la siguiente manera.

Diseño de grupo de control sólo con postest



Este diseño básico puede ampliarse para incluir varios tratamientos:



El diseño de grupo de control sólo con postest se utiliza cuando no es posible o adecuado realizar un pretest y en situaciones en las que el pretest podría tener un efecto en el tratamiento. Existen cuatro inconvenientes en el uso de un diseño sólo con postest en vez de un diseño pretest-postest: (1) si hay alguna posibilidad de que la asignación aleatoria no haya controlado las diferencias iniciales entre grupos, la ausencia de un pretest dificulta bien la comprobación de la existencia de las diferencias, bien el control estadístico de las diferencias que puedan encontrarse; (2) el investigador no puede formar subgrupos sobre la base del pretest para investigar los efectos del tratamiento en subgrupos diferentes; (3) el investigador no es capaz de determinar si se ha producido mortalidad experimental; y (4) el análisis estadístico es menos preciso, siendo menos probable que muestre una diferencia entre los grupos experimental y de control como en el diseño de pretest-postest. Si existen las siguientes condiciones, entonces el diseño pretest-postest sería entonces preferible:

1. Hay pequeñas diferencias entre las condiciones de tratamiento.
2. La mortalidad experimental es posible.
3. Se busca el análisis de subgrupos.
4. El anonimato no es necesario.
5. El pretest es una parte normal de la rutina de los sujetos.

EXTRACTO 8.1

DISEÑO DE GRUPO DE CONTROL CON PRETEST-POSTEST

El propósito del presente estudio era ampliar la investigación presentada anteriormente, considerando el efecto de comentarios valorativos de profesores realizados individualmente durante un periodo de tiempo determinado, en la percepción de control entre los alumnos más jóvenes del instituto de la ciudad. Las limitaciones de los estudios anteriores asociados al uso de grupos intactos fueron evitadas formando aleatoriamente los grupos experimental y de control. Además, los tratamientos fueron diseñados de modo que representaran intervenciones naturalistas en un programa de instrucción continuado, antes que un tratamiento «de efecto inmediato» que representara prácticas educativas discutibles.

Método

A los 196 alumnos se les administró el cuestionario Bialer durante su clase de ciencias. Bialer (1) diseñó esta escala de locus de control para ser usada con niños. Contiene

23 cuestiones del tipo: «¿crees que un chico puede ser lo que él quiera?» y el niño expresaba su acuerdo o desacuerdo con cada pregunta rodeando con un círculo la palabra sí o no. Hay dieciocho preguntas para las que un «sí» se puntúa como una respuesta interna y cinco cuestiones para las que un «no» se puntúa como una respuesta interna. La puntuación total es la suma del número de las respuestas internas señaladas. Se emplearon instrucciones modelo, excepto que el segundo autor leía las instrucciones y cada cuestión en voz alta a los alumnos mientras éstos la leían en silencio. Lefcourt (6) presenta las instrucciones, así como un listado completo de las cuestiones con sólo unos cuantos modelos de cuestiones que aparecen en el artículo de Bialer citado anteriormente. Sin embargo, ninguna fuente presenta datos con respecto a la fiabilidad de la escala. Las preguntas eran registradas directamente en el formulario del cuestionario y, más tarde, codificadas para los análisis de ordenador. Los cuestionarios del pretest estaban ordenados por niveles de puntuación y, dentro de cada rango, los sujetos fueron asignados a los grupos experimental o de control.

Los cinco profesores de las asignaturas principales, entre los que estaba el segundo autor, cooperaron en la administración de las condiciones de tratamiento a los 196 alumnos. A cada profesor se le facilitaron los nombres de los alumnos en cada condición del tratamiento. Se pidió a los profesores que escribieran comentarios alentadores personalizados en todos los trabajos y exámenes a los que se sometió a los alumnos en los grupos experimentales. Se les dio a los profesores las siguientes directrices: «los comentarios deben ser declaraciones positivas con independencia de la nota, que sean claramente visibles para el receptor y contengan el pronombre “tú” siempre que sea posible, como, por ejemplo, “tú estás mejorando”». Se pidió a los profesores que no escribieran ningún comentario en los trabajos o exámenes dirigidos a los alumnos del grupo de control salvo para la tradicional puntuación o nota. Las condiciones del tratamiento se impusieron durante un periodo de seis semanas, tras lo cual se volvió a administrar la prueba de locus de control de Bialer.

Fuente: Cross, L. H. y Cross, G. M. (1981), «Teachers evaluative comments and pupil of control», *Journal of Experimental Education*, 49, 68-71. Reimpreso con autorización de Helen Dwight Reid Educational Foundation. Publicado por Heldref Publications, 4000 Albermale St., N.W. Washington DC 20016. Copyright 1981.

Las ventajas del diseño sólo con posttest radican en que permite la evidencia experimental cuando es imposible realizar un pretest, impide el efecto reactivo del pretest y su uso facilita la garantía del anonimato. Por lo tanto, este diseño es especialmente bueno para la investigación de las actitudes por dos razones: el uso de un cuestionario de actitud como pretest podría afectar al estudio y, en general, las actitudes se presentan más honestamente cuando se garantiza el anonimato.

El diseño de grupo de control sólo con posttest controla prácticamente las mismas fuentes de invalidez que el diseño de pretest-posttest. La tabla 8.4 resume las fuentes de invalidez.

En el extracto 8.2 se emplea un diseño sólo con postest para investigar el efecto de varias estrategias de combinación de frases en habilidades de escritura, automotivación e interés intrínseco en la escritura. Había seis condiciones experimentales y un grupo de control.

EXTRACTO 8.2

DISEÑO DE GRUPO DE CONTROL SÓLO CON POSTEST

Designamos aleatoriamente a los 84 participantes a una de las seis condiciones experimentales o a un grupo de control sólo de práctica, con 12 chicas en cada grupo. Las condiciones experimentales estaban basadas en tres tipos de consecución de metas (proceso, logro y proceso-logro) y dos variaciones en el autoregistro (presente o ausente)... Como comprobación de la efectividad de la asignación aleatoria, comparamos el uso del inglés en el NEDT de las chicas en los seis grupos experimentales como línea base y no encontramos diferencias significativas... todas las chicas fueron sometidas a un postest en atribuciones, autoeficacia, habilidades de escritura, autorreacción e interés intrínseco. El experimentador comenzó y terminó cada sección del estudio y registró los resultados del postest.

Fuente: Zimmerman, B. J. y Kitsantas, A. (1999), «Acquiring Writing Revision skill: shifting from process to outcome self-regulatory goals», *Journal of Educational Psychology*, 91 (2), 244-245.

Ocasionalmente, surge la contingencia de que el investigador necesite descartar el efecto del pretest en el tratamiento. El diseño utilizado para este propósito es una combinación del grupo de control sólo con postest y el diseño de grupo de control con pretest-postest, y se llama el *diseño Solomon de cuatro grupos*. Aunque este diseño controla los efectos de la mortalidad y las interacciones pretest-tratamiento, es difícil aplicar el diseño en educación porque requiere el doble de sujetos y grupos que otros diseños.

DISEÑOS SEMIEXPERIMENTALES

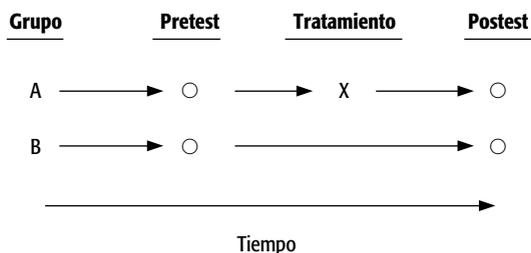
Los diseños experimentales «puros» proporcionan los argumentos más firmes y convincentes del efecto causal de la variable independiente porque controlan la mayoría de las fuentes de la invalidez interna. Sin embargo, hay muchas circunstancias en investigación educativa para las que, aunque la inferencia causal es deseable, no es posible diseñar experimentos auténticos, o en las que la necesidad de una fuerte validez externa es mayor. Las razones más comunes por las que no pueden emplearse diseños experimentales son que la asignación aleatoria de sujetos a los grupos experimental y de control resulte imposible y que un grupo de control o de comparación sea inconveniente, de-

masiado caro o no se pueda conseguir. Afortunadamente, existen varios buenos diseños que pueden utilizarse en cualquiera de estas circunstancias. Son denominados **diseños semiexperimentales** porque, aunque no son experimentos puros, ejercen un control considerable sobre la mayoría de las fuentes de invalidez y son, normalmente, más rigurosos que los diseños preexperimentales. Aunque hay muchos diseños semiexperimentales (Cook y Campbell, 1979), discutiremos sólo los más comunes.

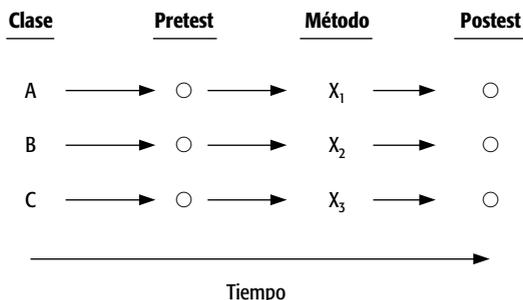
Diseño de grupos no equivalentes con pretest-posttest

Este diseño es muy habitual y útil en educación, dado que a menudo es imposible asignar sujetos de manera aleatoria. El investigador utiliza grupos de sujetos intactos previamente establecidos, pasa un pretest, administra la condición de tratamiento a cada grupo y pasa el posttest. Por lo tanto, la única diferencia entre este diseño y el diseño de grupo de control con pretest-posttest reside en la ausencia de asignación aleatoria de sujetos. Este diseño aparece representado a continuación:

Diseño de grupos no equivalentes con pretest-posttest



Como aparece en la tabla 8.5 de la página 344, la amenaza más seria a la validez interna de la investigación realizada con este diseño es la selección; es decir, como los grupos pueden diferir en las características que afectan a la variable dependiente, el investigador debe realizar la selección y dar argumentos razonables que expliquen que esta amenaza no es una hipótesis alternativa probable. Supongamos que un investigador está interesado en estudiar el efecto de tres métodos diferentes para modificar las actitudes de los profesores en formación hacia la instrucción asistida por ordenador. El investigador tiene tres clases de profesores con los que trabajar y es imposible asignar alumnos, de manera aleatoria, de cada clase a cada uno de los tres métodos. Por lo tanto, el investigador utiliza cada clase intacta y da a cada clase un tratamiento diferente. El diseño sería el siguiente:



La interpretación de los resultados dependerá en gran medida de si los grupos difieren en algunas características que podrían estar razonablemente relacionadas con la variable independiente. Se toma esta decisión comparando características de los tres grupos: sexo, tiempo que lleva el grupo formado, tamaño de los grupos, rendimiento, aptitud, estatus socioeconómico y resultados del pretest. Si, por ejemplo, la clase A comprende todos los alumnos de enseñanza primaria y las clases B y C los de secundaria, y los resultados demostraron que la clase A mejoró más que la B y la C, se podría atribuir este progreso a la diferencia entre los valores y a la formación de los alumnos de primaria y secundaria. Por otro lado, si las clases son casi idénticas en la mayoría de las características, entonces sería razonable asumir que las diferencias de la selección, probablemente no tendrán peso sobre los resultados. Por consiguiente, si el investigador sabe de antemano que la asignación aleatoria es imposible, los grupos deberían ser seleccionados de modo que fueran lo más parecidos posible. A continuación, se utilizan los resultados del pretest y otras medidas para ajustar los grupos estadísticamente en el factor evaluado. Otro enfoque para controlar la selección, cuando deben usarse grupos intactos como las clases, es trabajar con un gran número de grupos y, a continuación, asignar aleatoriamente grupos enteros a la condición experimental o de control. Este procedimiento modifica entonces el estudio por un diseño experimental. De hecho, éste es el enfoque preferente cuando la difusión del estudio o la historia son amenazas viables.

Las amenazas de maduración y la regresión estadística son las únicas otras diferencias entre este diseño y el diseño de grupo de control con pretest-postest. La regresión supone un problema si uno de los grupos obtiene resultados extremadamente altos o bajos. Por ejemplo, si un estudio para evaluar el impacto de un programa sobre niños superdotados seleccionó niños superdotados que tenían puntuaciones bajas y niños normales con resultados altos como grupo de comparación, entonces la regresión estadística hará que los resultados muestren una diferencia en las puntuaciones del postest cuando, en realidad nada ha cambiado. Los efectos de la maduración (aumentar la experiencia, cansancio, aburrimiento, etc.) dependerán de las diferencias específicas de las características entre los grupos.

En el extracto 8.3 se emplea un diseño semiexperimental de grupos no equivalentes con pretest-postest utilizando como sujetos a alumnos de clases intactas como sujetos. Observe cómo, aun existiendo una asignación aleatoria de clases, los investigadores indican con propiedad que no se trata de un diseño experimental puro porque los alumnos no fueron asignados aleatoriamente a las clases. El estudio tiene una única variable independiente con dos niveles y medidas de rendimiento y actitud como variables dependientes.

Diseños de series temporales

En el diseño de un grupo con pretest-postest, un único grupo de sujetos recibe un pretest y un postest. Si el grupo es medido en repetidas ocasiones antes y después del tratamiento, en lugar de una vez con anterioridad y otra con posterioridad, se crea un diseño diferente llamado «de serie temporal». Los **diseños de series temporales** son especialmente útiles cuando hay observaciones continuas que ocurren naturalmente sobre la variable dependiente y hay un tratamiento inesperado y distinto durante las observaciones. Estos diseños ofrecen mejoras significativas con respecto al diseño pretest-postest porque con una serie de preobservaciones y postobservaciones, los patrones de estabilidad y cambio pueden ser evaluados con más precisión. Hay muchos tipos diferentes de diseños temporales (Cook y Campbell, 1979); presentaremos los más comunes.

El diseño de series temporales interrumpidas con un grupo. Este diseño requiere un grupo y múltiples observaciones o evaluaciones antes y después del tratamiento. Las observaciones anteriores al tratamiento pueden ser consideradas como pretests repetidos y las posteriores, como postests. El diseño puede representarse en el siguiente diagrama.

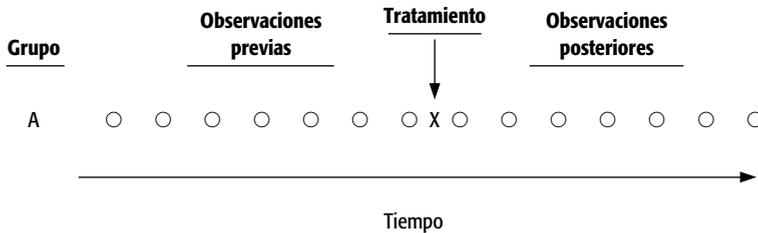
EXTRACTO 8.3

DISEÑO CUASIEXPERIMENTAL DE GRUPO DE COMPARACIÓN CON PRETEST-POSTEST

Se adoptó un diseño semiexperimental de grupo de control no equivalente (Campbell y Stanley, 1996) que implicaba 16 clases intactas. La asignación aleatoria de alumnos a nuevas clases no es probable en el sistema educativo de Taiwan; la clase intacta era la unidad del diseño experimental. Se asignaron aleatoriamente 8 clases intactas ($n = 319$) a la instrucción de preguntas en grupo y ocho clases ($n = 293$) al grupo de clase magistral tradicional. Se examinó y encuestó a los participantes de ambos grupos antes y después de la intervención de cuatro semanas.

Fuente: Chang, C. y Mao, S. (1999), «Comparison of Taiwan science student's outcomes with inquiry-group versus traditional instruction», *The Journal of Educational Research*, 96 (6), 342.

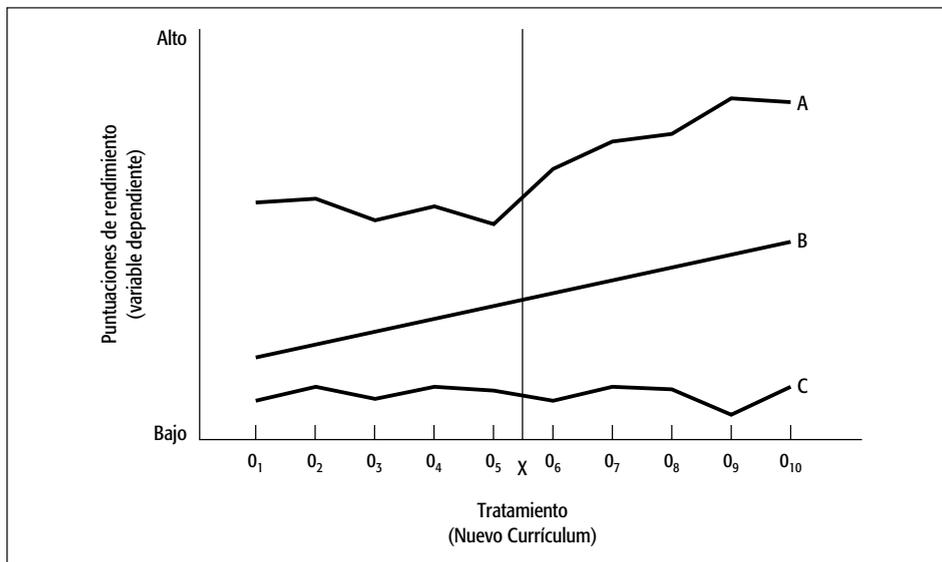
Diseño de series temporales interrumpidas con un grupo



Deberían darse varias condiciones al emplear este diseño. En primer lugar, las observaciones deberían hacerse en intervalos de tiempo iguales, y deberían realizarse siguiendo los mismos procedimientos para reducir la amenaza de la instrumentación. En segundo lugar, el estudio introducido debería ser una intervención clara y diferenciable que resulte completamente nueva en el contexto existente. En tercer lugar, debería existir alguna evidencia de que los sujetos implicados en cada observación son los mismos (es decir, que existe una baja mortalidad experimental). Una variación en la utilización de los mismos sujetos para cada medida consiste en emplear grupos diferentes pero muy parecidos. De este modo, por ejemplo, se podría evaluar muy bien un nuevo currículum. Por ejemplo, el rendimiento de los alumnos de sexto curso podría determinarse durante muchos años por medio del antiguo currículum y, a continuación, se podrían registrar durante varios años las puntuaciones de rendimiento después de la introducción del nuevo currículum. El elemento clave de este diseño es que las características de los alumnos de sexto curso deben ser aproximadamente las mismas año tras año. Obviamente, si con el paso de los años se produce el ingreso de alumnos más brillantes, el rendimiento aumentará independientemente del currículum.

En la figura 8.1 aparecen indicados algunos posibles resultados para el estudio. Si se consigue el resultado A, entonces, el investigador puede llegar a la conclusión de que el currículum tuvo un efecto positivo en el rendimiento. El resultado B indica una mejora fija de los resultados, por lo que es difícil interpretar el efecto del currículum, y el resultado C indica pocos cambios en el espacio de tiempo. Sin embargo, al interpretar estos resultados el investigador debería buscar explicaciones alternativas. Si parece que ha tenido lugar una modificación entre la población estudiantil como, por ejemplo, la migración desde el centro hasta los barrios, entonces se esperaría que las observaciones cambiaran. Se necesitaría que el instrumento de prueba fuera el mismo (no debería existir ningún cambio en el grupo normativo). Quizás, la más seria amenaza a la validez sea la historia. Es posible que otros hechos aparte del tratamiento, en este caso el currículum, ocurrieran al mismo tiempo y afectaran a las observaciones de posttest; por ejemplo, puede ser que en el mismo año en el que se cambió el currículum, los profesores también cambiaran. Otras amenazas incluyen la

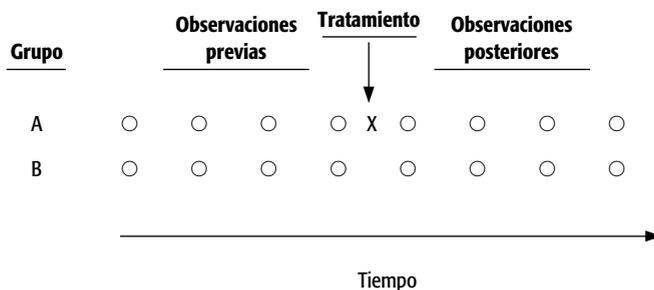
FIGURA 8.1: POSIBLES PATRONES DE RESULTADOS DE RENDIMIENTO A TRAVÉS DEL TIEMPO PARA DISEÑOS DE SERIES TEMPORALES



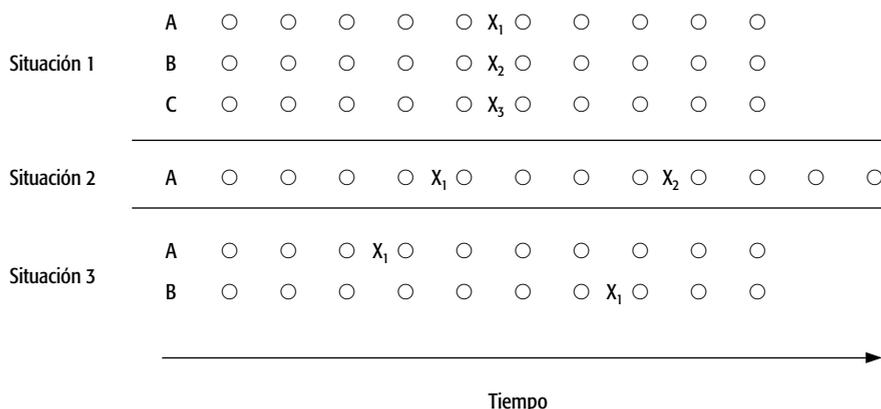
variación estacional (las puntuaciones de autoestima pueden ser más bajas en invierno que en primavera) y el pretest (el efecto del pretest en el estudio).

El diseño de series temporales interrumpidas con grupo de control. En este diseño se añade un grupo de control o de comparación al de series temporales interrumpidas con un grupo. La introducción de un grupo de control fortalece el diseño considerablemente, ya que así se elimina la amenaza principal de la historia. La instrumentación es también una explicación menos probable y, si se incluye la asignación aleatoria, entonces la selección no supone una amenaza para la validez. Sin embargo, a partir del momento en que un grupo de control está presente, la difusión del estudio se convierte en una amenaza. Este diseño aparece representado a continuación.

Diseño de series temporales interrumpidas con grupo de control



Existen muchas variantes de los diseños de series temporales básicos. Por ejemplo, se puede eliminar, en lugar de añadir, un tratamiento o bien se pueden comparar múltiples tratamientos ya sea con uno o varios grupos. En el siguiente diagrama se indican algunas variantes. En la situación 1 se comparan tres diferentes tratamientos, usando tres grupos de sujetos. En la situación 2 sólo se utiliza un grupo de sujetos y se comparan dos tratamientos y en la situación 3 se compara el mismo tratamiento en dos grupos y en diferentes momentos en el tiempo.



Los diseños semiexperimentales que se han presentado en este capítulo son diseños simples y básicos que, con frecuencia, se amplían en estudios reales y existen muchos diseños que no han sido mencionados. La elección del diseño dependerá de las variables estudiadas, de las circunstancias del contexto en que se realiza la investigación y de la posibilidad de amenazas a la validez interna. El aspecto importante es que existen limitaciones en todos los diseños de investigación y es necesario para el investigador y para el lector de investigación averiguar y analizar las posibles hipótesis alternativas que podrían explicar los resultados. La tabla 8.5 resume las amenazas a la validez interna de los diseños semiexperimentales.

DISEÑOS DE CASO ÚNICO

Los diseños preexperimentales, experimentales y semiexperimentales que han sido descritos están basados en un concepto tradicional de investigación según el cual el comportamiento se investiga mejor recurriendo a grupos de sujetos. Normalmente, por ejemplo, estamos interesados, en las actitudes, en general, de alumnos de cuarto curso, o en si, en general, un método particular de lectura es mejor. Sin embargo, existen muchas circunstancias en las que o bien no se quiere o bien es imposible emplear grupos de sujetos, como cuando se examinan estrategias de instrucción para aplicarlas con alumnos particulares. En estas

TABLA 8.5: AMENAZAS A LA VALIDEZ INTERNA DE LOS DISEÑOS SEMIEXPERIMENTALES

Diseño	Historia	Selección	Regresión estadística	Pretest	Instrumentación	Mortalidad experimental	Maduración	Difusión del tratamiento	Efectos del experimentador	Repeticiones del tratamiento	Efectos del sujeto	Conclusión estadística
Diseño de grupos no equivalentes con pretest-postest	?	-	?	+	?	?	-	?	?	?	?	?
Series temporales interrumpidas con un grupo	-	?	+	?	?	?	+	SA	+	?	?	?
Series temporales interrumpidas con grupo de control	+	?	+	+	?	?	+	?	?	?	?	?

Observación: en esta tabla, así como en las tablas 8.3 y 8.4 un signo menos significa una limitación definida; un signo más significa que el factor está controlado; una interrogación significa una posible fuente de invalidez y SA indica que la amenaza no es aplicable a este diseño (y, por lo tanto, no es tampoco un factor).

situaciones, **los diseños de caso único** se emplean, a menudo, para proporcionar deducciones causales rigurosas sobre el comportamiento de uno, dos o unos pocos individuos. En realidad, el término «caso único» se refiere al modo en que se presentan y analizan las puntuaciones, por sujetos particulares, en contraste con los diseños de grupo que utilizan resultados promedio. El enfoque básico es estudiar individuos en una condición de no tratamiento y, a continuación, en una condición de tratamiento; evaluando de forma continua en ambos casos la actuación en la variable dependiente.

Las características del diseño que consiguen una elevada validez interna con diseños de un único sujeto son, de algún modo, diferentes a las técnicas tratadas anteriormente en el contexto de los diseños de grupo. Las características más importantes de los diseños de caso único están resumidas a continuación:

1. **Medición fiable.** Los diseños de caso único, normalmente, implican muchas observaciones de la conducta como técnica de recogida de datos. Es importante que se estandaricen las condiciones de observación, tales como momento del día y lu-

gar; que los observadores están bien entrenados y se comprueben su fiabilidad y sesgos y que se defina operacionalmente la conducta a observar. La consistencia en la medida es especialmente importante cuando el estudio va de una condición a otra. Dado que una medida precisa es crucial para los diseños de caso único, normalmente el investigador informa sobre todos los aspectos de la recogida de datos para que se pueda descartar razonablemente cualquier amenaza a la validez.

2. **Medidas repetidas.** Una característica distintiva de los diseños de caso único es la medición, repetidas veces, de un mismo aspecto de la conducta, y del mismo modo a lo largo del estudio. Esto es muy diferente de la medición en muchos estudios de grupo en los que hay una única medida antes o después del tratamiento. La medición repetida controla la variación normal que cabe esperar en cortos periodos de tiempo y ofrece una descripción clara y fiable de la conducta.

3. **Descripción de las condiciones.** Debería facilitarse una descripción precisa y detallada de todas las condiciones en las que se observa la conducta. Esta descripción permite la aplicación del estudio a otros individuos para fortalecer tanto la validez interna como la externa.

4. **Línea base y condición de tratamiento; duración y estabilidad.** El procedimiento habitual es que cada condición dure más o menos la misma cantidad de tiempo y contenga aproximadamente el mismo número de observaciones. Si varían la cantidad de tiempo o el número de observaciones, entonces tiempo y número de observaciones se convierten en variables de confusión que complican la interpretación de los resultados y debilitan la validez interna. También es importante que se observe el comportamiento el tiempo necesario para establecer un patrón fijo. Si hay variaciones considerables en la conducta, entonces será difícil determinar si los cambios observados se deben a la variación natural o al tratamiento. Durante la primera fase de investigación de caso único se observa la conducta objeto de estudio bajo condiciones naturales hasta que se consigue la estabilidad. Este periodo de tiempo se llama **línea base**. La fase de tratamiento tiene lugar con un cambio en las condiciones por parte del investigador y debe ser también lo suficientemente larga como para conseguir la estabilidad.

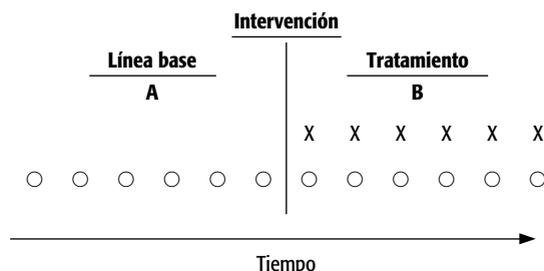
5. **Regla de variar de uno en uno.** Es importante cambiar sólo una variable durante la fase de tratamiento en la investigación de caso único y la variable cambiada debe describirse con precisión. Si se cambian dos o más variables al mismo tiempo, el investigador no puede estar seguro de qué cambio o cambios provocaron los resultados.

Diseños A-B

Para distinguir los diseños de caso único de los diseños tradicionales de grupo, se utiliza una convención específica de notación. En ésta, las letras en vez de representar grupos de sujetos, se presentan como condiciones: A se presenta como la condición de la línea base y B como la condición del tratamiento.

El **diseño A-B** es el diseño de caso único más simple y menos interpretable. El procedimiento consiste en observar la conducta objetivo hasta que se da en una tasa fija y consistente. Esta condición es la línea base o condición A. Entonces, se introduce un tratamiento en el entorno en el que se han recogido los datos de la línea base y a esta condición se la llama B. El diseño puede representarse con el siguiente diagrama:

Diseño de caso único A-B

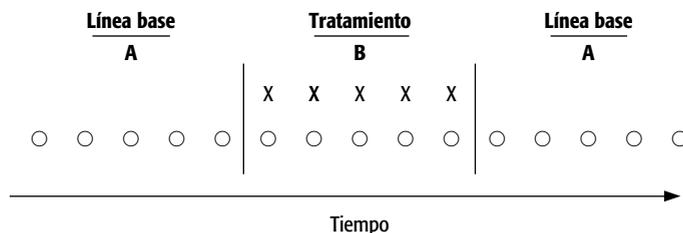


La interpretación de los resultados se basa en la premisa de que si no se introdujera ningún tratamiento, el comportamiento continuaría tal como se registró en la línea base. Si el comportamiento cambia durante la condición de tratamiento, este cambio puede atribuirse a la intervención introducida por el investigador. Sin embargo, a menudo, otros factores como la administración de pruebas y la historia no pueden descartarse razonablemente en este diseño, de manera que es relativamente débil en validez interna.

Diseños A-B-A

El **diseño A-B-A**, también denominado **diseño de reversión o de reserva** es un diseño más común en la investigación de caso único, en el que se añade un segundo periodo de línea base tras el tratamiento. En este diseño, que aparece representado a continuación, el investigador establece una línea base (A), introduce el tratamiento (B) y, a continuación, elimina el tratamiento para restablecer la condición (A) de línea base.

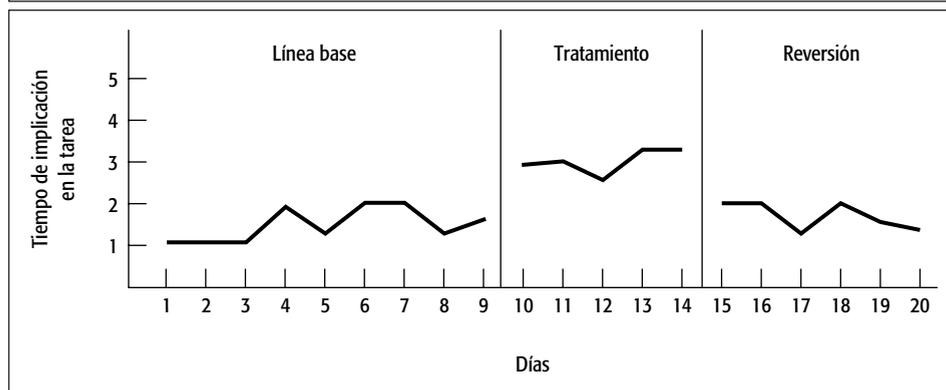
Diseño de caso único A-B-A



Este diseño permite la realización de deducciones causales rigurosas si el patrón de conducta cambia durante la fase de tratamiento y, después, una vez que éste se ha eliminado vuelve prácticamente al mismo patrón y a ser observado en la primera línea base. Como ejemplo hipotético para ilustrar este diseño, supongamos que un profesor está interesado en probar una nueva técnica de refuerzo con John, uno de los alumnos de quinto curso, con la esperanza de que la nueva técnica aumente el tiempo que realmente John pasa estudiando (tiempo de implicación en la tarea). El profesor primero registra la cantidad de tiempo de estudio aproximado diario hasta que se establece un patrón estable. Entonces, el investigador introduce la técnica de refuerzo como intervención y sigue observando el tiempo en la tarea. Tras un espacio de tiempo determinado, el profesor deja de usar la técnica de refuerzo para ver si la conducta durante la tarea vuelve a la condición de la línea base. La figura 8.2 ilustra los resultados de este estudio hipotético que ofrecen una clara evidencia de la existencia de un vínculo causal entre la técnica de refuerzo y el aumento del tiempo de estudio.

Se puede obtener una mayor evidencia de un cambio en el comportamiento causado por el tratamiento si el diseño A-B-A se amplía hasta reinstaurar el tratamiento, convirtiéndose en A-B-A-B. El diseño A-B-A-B no sólo ofrece una mayor inferencia causal que el diseño A-B-A, sino que, además, finaliza con la condición del tratamiento lo que con frecuencia, por razones éticas, es más favorable para el sujeto. Si el patrón de resultados no consigue apoyar el efecto del tratamiento, entonces, la interpretación es menos clara. Si cambia el comportamiento durante el tratamiento, pero no logra volver a la condición de línea base una vez finalizado el tratamiento, el investigador no sabrá si intervienen otros factores aparte del tratamiento que causaron el cambio o si el tratamiento había sido tan efectivo que podía ser eliminado y, sin embargo, provocar un impacto en el comportamiento.

FIGURA 8.2: RESULTADOS DE UN ESTUDIO HIPOTÉTICO CON EL USO DE UN DISEÑO A-B-A



El diseño A-B-A-B aparece ilustrado en el extracto 8.4. El propósito de este experimento era investigar los efectos de las reprimendas dirigidas a un estudiante sobre un segundo alumno.

EXTRACTO 8.4

DISEÑO DE CASO ÚNICO A-B-A-B

Diseño experimental

En este experimento se usó un diseño de reversión. Tras obtener una línea base estable de ejecución tanto para Jeanette como para Natalie, se reprendió el comportamiento revoltoso de Jeanette siguiendo un esquema de castigo de intervalo variable de dos minutos. Durante esta condición, no se criticó el comportamiento revoltoso de Natalie. Tras volver a las condiciones de la línea base, se hicieron eventualmente reprimendas a las conductas disruptivas de Jeanette. Tras este tratamiento, se terminó el experimento porque el periodo escolar había concluido.

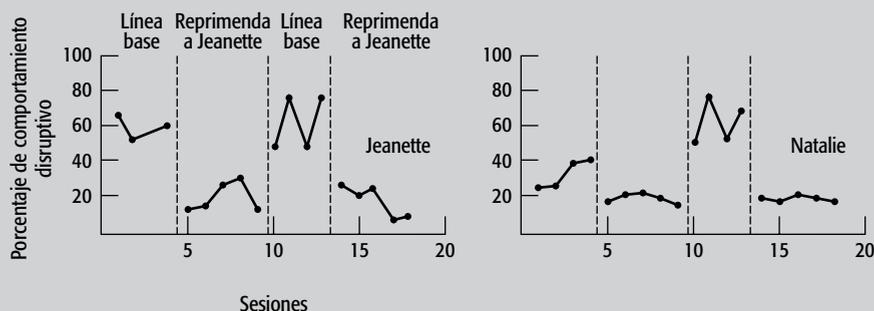
Línea base 1. Durante esta condición se dieron órdenes al profesor de no elogiar ni reprender a Jeanette o a Natalie y dar la clase como de costumbre.

Reprimenda a Jeanette 1. Durante esta condición el profesor reprendió el comportamiento revoltoso de Jeanette con un programa de castigo de intervalo variable empleando intervalos de 2 minutos. Al profesor se le indicó cuándo hacer las reprimendas utilizando el mismo procedimiento que en el experimento precedente. Al igual que en el experimento 3A, se utilizaron varias indicaciones.

Línea base 2. Esta condición se trató del mismo modo que la primera condición de la línea base.

Reprimenda de Jeanette 2. Esta condición se realizó del mismo modo que la primera condición de reprimenda de Jeanette.

Porcentaje de intervalos de 10 segundos en los que Jeanette y Natalie tenían un comportamiento disruptivo durante cada sesión del experimento.



Fuente: Van Houton, R. y otros, (1982), «An analysis of some variables influencing the effectiveness of reprimands», *Journal of Applied Behavior Analysis*, 15, 65-83. Copyright 1982 de la Society for the Experimental Analysis of Behavior, reimpresso con autorización.

Diseños de línea base múltiple

Cuando no se quiere o es imposible eliminar una condición de tratamiento, o cuando los efectos de una condición de tratamiento se extienden a una segunda fase de línea base, pueden hacerse firmes deducciones causales utilizando **diseños de línea base múltiple** en vez de un diseño simple A-B. Los diseños de línea base múltiple utilizan la lógica A-B, pero en vez de emplear un sujeto y un tipo de conducta objetivo, el investigador recoge información sobre dos o más acciones, sujetos, o situaciones o alguna combinación de acciones, situaciones y sujetos.

Líneas base múltiples entre comportamientos. En este diseño se registran en un sujeto medidas de línea base sobre dos o más conductas discretas e independientes. Tras haberse establecido una línea base estable para todas las conductas, el tratamiento se aplica primero a un comportamiento y, a continuación, tras un intervalo constante de tiempo, se aplica al segundo y así hasta que todos hayan sido sometidos al tratamiento.

Puede hacerse una inferencia causal a partir del efecto del tratamiento si la actuación muestra un cambio consistente sólo tras haberse introducido el tratamiento para cada tipo de comportamiento. Para facilitar una comparación significativa es necesario empezar los tratamientos en diferentes momentos para cada uno. De este modo, el comportamiento que permanece en la condición de línea base controla los que están recibiendo la condición de tratamiento. El problema más preocupante de este diseño es el uso de dos o más comportamientos tan parecidos que la primera vez que se introduce el tratamiento éste afecta a ambos.

Líneas base múltiples entre situaciones. En este diseño se somete a observación, en dos o más escenarios, un tipo único de comportamiento de un individuo. Por ejemplo, un profesor podría estar interesado en investigar si un estudiante respondería del mismo modo a un *feedback* individualizado en matemáticas, ciencias e inglés. El diseño básico es el mismo que en el diseño de la línea base múltiple entre comportamientos, salvo que la situación reemplaza los tipos de comportamiento como la condición que varía (como, por ejemplo, aprender a comportarse tanto en una clase como en una tienda de comestibles o en una clase y en la cafetería).

Líneas base múltiple entre individuos. Este diseño utiliza dos o más individuos y mantiene constante la conducta y la situación.

Tras haberse observado una línea base estable para un sujeto, se introduce el tratamiento sólo para ese sujeto. Tras un intervalo determinado, el segundo sujeto recibe el tratamiento, y así sucesivamente.

Este diseño es efectivo en la medida en que los sujetos implicados no estén influenciados por otros que ya hayan recibido el tratamiento (por ejemplo, con

alumnos que se encuentren en la misma clase o hermanos). Un uso adecuado de este tipo de diseño sería hacer que un profesor empleara el mismo procedimiento de estudio con alumnos de cuatro clases diferentes.

En los tres diseños de línea base múltiple son posibles muchas variantes. Una buena fuente para una visión más detallada de estos diseños es Barlow y Hersen (1984), Kazdin (1982) y Franklin, Allison y Gorman (1977). Los formatos A-B-A y A-B-A-B pueden combinarse con diseños de línea base múltiple. Los diseños que implican reversión, retirada y restablecimiento del tratamiento son, por lo general, más rigurosos con respecto a la validez interna. Como se puede imaginar, la validez externa de los diseños de caso único es bastante limitada. La generalizabilidad de los resultados de un estudio aumenta, principalmente, mediante la repetición con otros sujetos y en situaciones diferentes.

NORMAS DE ADECUACIÓN

Al juzgar la adecuación de los diseños presentados en este capítulo debería centrar su atención en algunos criterios clave. Estos criterios aparecen enumerados aquí en forma de preguntas que deberían plantearse para cada tipo de diseño.

Diseños experimentales

1. ¿Estaba el diseño de la investigación descrito con suficiente detalle como para permitir una replica del estudio?
2. ¿Estaba claro cómo se conseguía la equivalencia estadística de los grupos? ¿Había una descripción completa de la manera específica en la que se designó aleatoriamente los sujetos a los grupos?
3. ¿Era apropiado para el problema de investigación un diseño experimental?
4. ¿Hubo manipulación de la variable independiente?
5. ¿Hubo un control máximo sobre las variables extrañas y los errores de medida?
6. ¿Era la condición de tratamiento lo suficientemente diferente de la condición de comparación como para que se pudiera esperar un efecto diferencial sobre la variable dependiente?
7. ¿Fueron razonablemente descartadas o señaladas y discutidas las amenazas potenciales a la validez interna?
8. ¿Se describió el intervalo temporal del estudio?
9. ¿Evitó el diseño ser demasiado artificial o restringido para conseguir una validez externa adecuada?
10. ¿Se consiguió un equilibrio apropiado entre el control de variables y las condiciones naturales?
11. ¿Se emplearon las pruebas de estadística inferencial apropiadas?

Diseños semiexperimentales

1. ¿Estaba el diseño de la investigación descrito con suficiente detalle como para permitir la réplica del estudio?
2. ¿Era posible un experimento?
3. ¿Estaba claro cómo se controlaban o descartaban como posibles hipótesis alternativas las variables extrañas?
4. ¿Se tuvieron en cuenta todas las amenazas potenciales a la validez interna?
5. ¿Descartaron las explicaciones, de manera razonable, las posibles hipótesis contrarias?
6. ¿Habría sido mejor un diseño semiexperimental diferente?
7. ¿Se acercó el diseño todo lo posible a un experimento puro?
8. ¿Había un equilibrio apropiado entre el control para la validez interna y la externa?
9. ¿Se hizo un esfuerzo en emplear grupos que fueran tan equivalentes como fuera posible?
10. Si se empleó un diseño de series temporales, (a) ¿Había un número adecuado de observaciones para sugerir un patrón de resultados? (b) ¿Se introdujo el tratamiento de forma distintiva en cada espacio de tiempo? (c) ¿La evaluación de la variable dependiente era consistente? (d) ¿Estaba claro, si se usaron grupos de comparación, cómo eran de equivalentes los grupos?

Diseños de caso único

1. ¿Era uno el tamaño muestral?
2. ¿El diseño de caso único era el más apropiado o habría sido mejor un diseño de grupo?
3. ¿Estaban estandarizadas las condiciones de observación?
4. ¿El comportamiento observado estaba definido de manera operacional?
5. ¿Era altamente fiable la medición?
6. ¿Había suficientes medidas repetidas?
7. ¿Se describieron completamente las condiciones en las que se realizó el tratamiento?
8. ¿Había estabilidad en la condición de la línea base antes de que se introdujera el tratamiento?
9. ¿Existía una diferencia entre el espacio de tiempo o el número de observaciones entre la línea base y las condiciones de tratamiento?
10. ¿Se cambió sólo una variable durante la condición de tratamiento?
11. ¿Se tuvieron en cuenta las amenazas a la validez externa e interna?

RESUMEN

El propósito de este capítulo ha sido introducir diseños que permitan la investigación del efecto causal de una variable sobre otra. El reto para la mayoría de los investigadores se encuentra en utilizar un diseño que, dadas las condiciones de la investigación, resulte el más adecuado para lograr su meta. Los principales aspectos del capítulo están resumidos a continuación.

1. La investigación experimental, tal como se define en las ciencias naturales, implica la manipulación de las variables experimentales y la asignación aleatoria de sujetos a grupos para investigar las relaciones de causa-efecto.
2. La investigación experimental clásica se caracteriza por la asignación aleatoria de sujetos a grupos de tratamiento y de control, la manipulación de las variables independientes y un estrecho control de las variables extrañas.
3. Un control estricto de las variables extrañas en la investigación experimental educativa podría limitar la generalizabilidad de los resultados.
4. La planificación de la investigación experimental implica la creación de grupos de comparación y experimentales, la manipulación del factor del grupo que recibirá el tratamiento y la evaluación del efecto del tratamiento en el comportamiento.
5. El elemento clave en la interpretación de los estudios experimentales es descartar las posibles hipótesis alternativas.
6. Los diseños preexperimentales controlan muy pocas amenazas a la validez interna.
7. Los diseños experimentales controlan la mayoría de las amenazas a la validez interna, pero algunas amenazas, tales como la historia y la difusión del tratamiento pueden constituir posibles hipótesis alternativas.
8. Los diseños semiexperimentales son empleados a menudo a causa de las dificultades que implica la realización de experimentos.
9. Los diseños de series temporales, en los que se hacen muchas observaciones antes y después del tratamiento, son especialmente útiles en los casos en los que la realización periódica de pruebas forma parte natural del contexto.
10. Los diseños de caso único aportan técnicas para hacer firmes deducciones causales sobre el efecto de un tratamiento sobre un individuo determinado o sobre un grupo.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La investigación experimental clásica, tal y como se realiza en las ciencias naturales, implica...
 - a. manipulación, asignación aleatoria, aplicación de resultados.
 - b. asignación aleatoria, manipulación, control.
 - c. control, manipulación, generalizabilidad.
 - d. generalizabilidad, control, independencia.
2. Una desventaja de la investigación experimental en entornos educativos es que...
 - a. es difícil realizar deducciones causales.
 - b. la validez interna es débil.
 - c. normalmente, su realización es cara.
 - d. a menudo, la generalizabilidad es limitada.
3. El diseño que tiene la validez interna más débil es...
 - a. el de un grupo sólo con postest.
 - b. el de un grupo con pretest-postest.
 - c. el diseño de caso único A-B-A-B.
 - d. los diseños de series temporales.
4. Con frecuencia, los resultados de los diseños preexperimentales tienen muchas posibles hipótesis alternativas a causa de...
 - a. una planificación insuficiente.
 - b. la falta de asignación aleatoria.
 - c. una falta de generalizabilidad.
 - d. el uso de un sólo grupo.
5. ¿Cuáles de las siguientes fuentes de validez interna controlan los diseños experimentales?
 - a. Selección, pretest.
 - b. Regresión, historia.
 - c. Selección, maduración.
 - d. Difusión del tratamiento, selección.
6. La ventaja más importante de utilizar experimentos, con respecto a la validez interna, es que se controla...
 - a. la selección.
 - b. la historia.
 - c. el pretest.
 - d. los efectos del experimentador.
7. Qué fuente de invalidez controla el diseño de grupos no equivalentes con pretest-postest...

- a. la regresión.
 - b. la historia.
 - c. el pretest.
 - d. la selección.
8. La característica distintiva de los diseños de series temporales es que...
- a. se comparan varios tratamientos.
 - b. se hacen muchas observaciones previas y posteriores.
 - c. la historia se controla con efectividad.
 - d. los sujetos son asignados aleatoriamente a grupos en momentos diferentes.
9. ¿A cuál de las siguientes amenazas a la validez interna es particularmente vulnerable el diseño de caso único A-B?
- a. Difusión del tratamiento, selección.
 - b. Selección, historia.
 - c. Difusión del tratamiento, regresión.
 - d. Instrumentación, historia.
10. Cuál de los siguientes conjuntos de características distingue los diseños de caso único de otros diseños experimentales?
- a. Una buena generalizabilidad, uso de datos de la línea base.
 - b. Uso de métodos múltiples, variación de una sola variable.
 - c. Uso de grupos de comparación, asignación aleatoria.
 - d. Uso de asignación aleatoria, variación de una sola variable.
11. ¿Para qué tipo de situación son especialmente útiles los diseños de línea base múltiple?
- a. Cuando están disponibles dos o más sujetos.
 - b. Cuando no es deseable usar sólo una línea de base.
 - c. Cuando no es deseable eliminar el tratamiento.
 - d. Cuando es posible estudiar dos o más tipos de conducta.

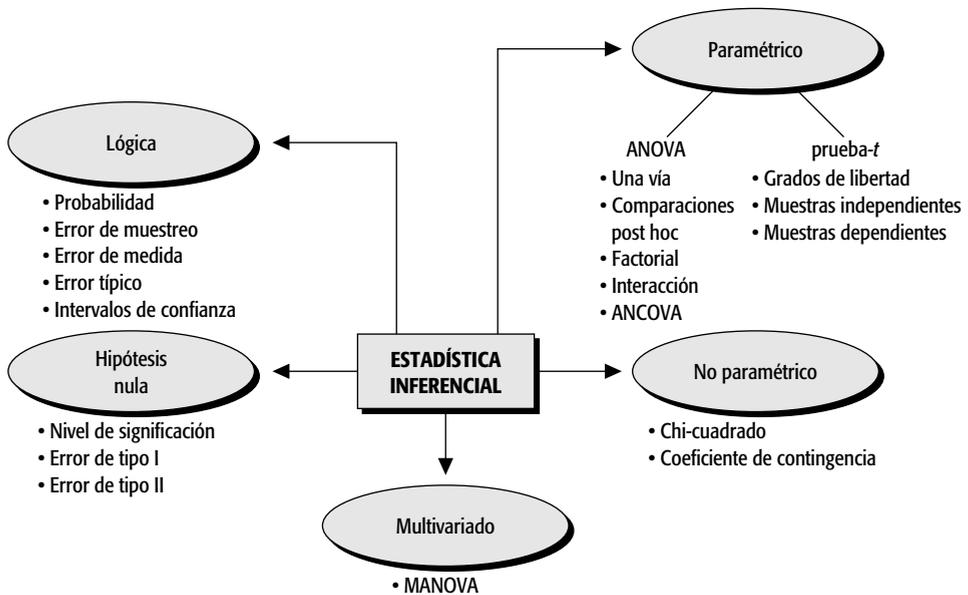
Problemas aplicados

Determine, en cada uno de los siguientes casos, el diseño utilizado y representelo utilizando el sistema de notación comentado en el capítulo.

1. Un investigador quiere probar la eficacia de tres métodos de enseñanza aplicándolos a un grupo de alumnos de undécimo curso. El investigador localiza un colegio que quiere cooperar y un profesor permite al investigador usar tres de sus clases. Entonces, el investigador administra un pretest a todos los alumnos, cada clase recibe un método de enseñanza diferente durante dos semanas y, después, el investigador pasa a todos los alumnos un postest.

2. Un profesor está interesado en determinar el efecto de emplear un sistema con los alumnos para controlar su comportamiento. El profesor decide registrar el comportamiento de los dos estudiantes, un chico y una chica, que parecen tener más problemas que el resto de los alumnos. Durante dos semanas el profesor registra su comportamiento. Los últimos días de la segunda semana el profesor empieza a usar el sistema con el chico y, al mismo tiempo, sigue registrando el comportamiento durante otras dos semanas. La chica no es sometida a tratamiento hasta los últimos días de la tercera semana.
3. Un investigador está interesado en saber si el orden de las preguntas de opción múltiple afecta al número de cuestiones respondidas correctamente. El investigador plantea tres tipos de examen: primero, una con las cuestiones fáciles en primer lugar y después las difíciles; otra con las cuestiones fáciles al final y las difíciles en primer lugar, y una tercera sin ningún orden, con cuestiones fáciles y difíciles mezcladas. Se administra la prueba a una clase de sesenta alumnos. Las pruebas están organizadas en veinte columnas, cada una de ellas con los tipos 1, 2 y 3. A continuación, se juntan las columnas y se reparten las pruebas entre los alumnos. Entonces, el investigador compara los resultados promedio de los alumnos que han respondido cada tipo de prueba.

Estadística inferencial



PALABRAS CLAVE

probabilidad	análisis de varianza (ANOVA)
distribución muestral	comparaciones <i>post hoc</i>
error típico	comparaciones planeadas
intervalo de confianza	ANOVA factorial
hipótesis estadística	interacción
hipótesis nula	análisis de covarianza (ANCOVA)
nivel de significación	paramétrico
error de tipo I	no paramétrico
error de tipo II	chi-cuadrado
nivel alfa	prueba de chi-cuadrado de muestras independientes
significativo estadísticamente	tabla de contingencia
tamaño del efecto	coeficiente phi
prueba t	coeficiente de contingencia
grados de libertad	multivariada
prueba t de muestras independientes	
prueba t de muestras dependientes	

Esta es la parte de los libros de investigación que asusta a la mayoría de los lectores. La expresión *Estadística inferencial* provoca ansiedad y miedo entre los alumnos ya ocupados en los llamados más simples procedimientos de análisis de datos descriptivos. Estos sentimientos se basan en la creencia errónea de que para entender la Estadística inferencial se necesitan cálculos matemáticos complejos y difíciles. Es cierto que los cálculos reales asociados a los procedimientos estadísticos inferenciales son complicados, pero no es necesario aprender ecuaciones y cálculos completos para entender y utilizar los resultados de estos procedimientos (en el Apéndice C se presentan algunos cálculos). Aprender los principios de la Estadística inferencial requiere estudio y aplicación, pero se trata más bien de comprender su lógica que de cálculos matemáticos.

En este capítulo presentaremos la lógica en la que se basa la Estadística inferencial, los principios de la comprobación de hipótesis y algunos procedimientos estadísticos comunes. Haremos hincapié en presentar los conceptos para poder entender y valorar el uso de estos procedimientos en la lectura y diseño de investigación.

LA LÓGICA DE LA ESTADÍSTICA INFERENCIAL

En algunos aspectos sería fantástico poder predecir los resultados de los acontecimientos. Cuando vamos a ver una película, ¿cómo podemos estar seguros de

que nos gustará? Cuando un profesor utiliza un procedimiento particular de agrupamiento, ¿qué seguridad tendrá de que funcionará? ¿Cuánta confianza tiene un paciente en que su operación será un éxito? ¿Están seguros los granjeros de que lloverá lo suficiente para sus cosechas? Las preguntas son interminables porque es propio de la naturaleza humana intentar predecir el futuro, y el grado de seguridad que tenemos sobre las predicciones varía enormemente. Existen muy pocas cosas en el mundo de las que podemos estar absolutamente seguros, y en ciencias sociales y en educación, normalmente, existe una cantidad considerable de incertidumbre. Por lo tanto, al hacer consideraciones sobre fenómenos investigados debemos emplear un lenguaje que refleje la naturaleza probabilística del caso. Los números, conceptos y términos empleados por la Estadística inferencial reflejan este lenguaje. Aunque se maneja un gran número de procedimientos de Estadística inferencial, muchos bastante complicados, el propósito es siempre el mismo. El objetivo es determinar con precisión la probabilidad de algo.

La probabilidad

La **probabilidad** es una manera científica de establecer el grado de confianza que tenemos para predecir algo. Kerlinger (1979) define la probabilidad de manera teórica como «el número de casos “favorables” de un acontecimiento dividido entre el número total de (igualmente posibles) casos» (p. 67-68). (*Favorable* aquí significa favorable a un acontecimiento cuya probabilidad estamos evaluando). Por lo tanto, si el número total de casos es doce, y el número de los casos favorables es 6, la probabilidad del caso favorable es de 6 entre 12 o .50. Los juegos de dados son un ejemplo más concreto. Un dado posee un total de seis posibles casos. Si el caso favorable obtiene un cuatro, la probabilidad de obtener realmente un cuatro sería de 1 entre 6, o .17; es decir, si se lanza un dado 100 veces, se obtendrán aproximadamente 17 cuatros. Si tenemos dos dados, ¿cuál es la probabilidad de obtener un siete? Como hay un total de 36 combinaciones diferentes, al lanzar dos dados (1 y 1; 1 y 2; 1 y 3 y así sucesivamente) y sólo 6 de obtener un 7 (1 y 6; 2 y 5; 3 y 4; 6 y 1; 5 y 2; y 4 y 3), la probabilidad de obtener un siete es de seis entre 36 o .17. ¿Y qué hay del lanzamiento de dos seises? Sólo hay una combinación de números, de modo que la probabilidad es de 1 entre 36 o .03.

Esta lógica se aplica en investigación a situaciones más complicadas. Por ejemplo, ¿cómo pueden estar seguros los que realizan un sondeo de que sus predicciones serán acertadas? ¿Cuál es la probabilidad de estar equivocados o de estar en lo cierto? Por ejemplo, en el caso de las encuestas Gallup, ¿cómo podemos estar seguros de que los resultados de una muestra denominada científica reflejan con precisión las actitudes del pueblo norteamericano? Cuando los profesores emplean el refuerzo positivo, ¿qué seguridad tienen de que aumentará la

frecuencia del comportamiento esperado? En estas situaciones el número de casos totales puede ser desconocido y la evaluación de los acontecimientos imperfecta. Hacemos así valoraciones de probabilidad influenciadas por la cantidad de error posible tanto en la medida como en el muestreo de acontecimientos.

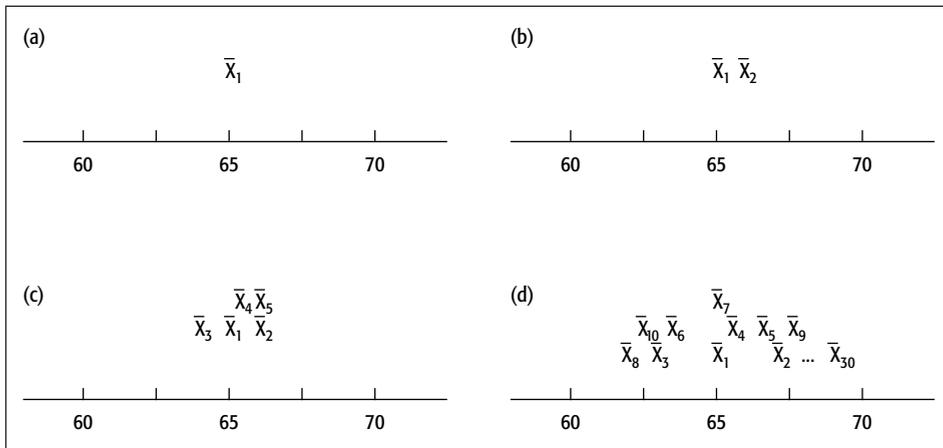
Error de muestreo y de medida

En el capítulo 4 se habló del muestreo como una técnica para estudiar a una parte de la población de todas las observaciones o los acontecimientos considerados. La población es el grupo más amplio al que el investigador pretende generalizar los resultados obtenidos de la muestra. Como subgrupo de la población, la muestra se usa para conseguir datos, y a continuación, la Estadística inferencial se emplea para hacer generalizaciones a la población. Imaginemos, por ejemplo, que un investigador está interesado en establecer la autoestima de los alumnos de cuarto curso en una escuela. El investigador podría evaluar la autoestima de todos los alumnos de cuarto, pero requeriría mucho tiempo, resultaría caro y probablemente innecesario. En vez de eso, el investigador toma una muestra de los alumnos de cuarto y mide la autoestima de cada uno. A continuación, el investigador *infiere* cuál es la autoestima de todos los alumnos de cuarto curso a partir de los resultados de la muestra seleccionada. La Estadística inferencial se utiliza para hacer estas deducciones. El grupo de todos los alumnos de cuarto curso es la población y el investigador emplea Estadística descriptiva (media y desviación típica) para estimar las características de la población.

¿Dónde interviene la probabilidad en este proceso? Cuando se describe una muestra, los resultados estadísticos representan una estimación imperfecta de la población. Hay un error en la obtención de la muestra y la probabilidad está relacionada con nuestra confianza en que el hecho de que la muestra representa con precisión la población. Aunque el investigador maneje una muestra aleatoria, la media y la varianza de esa muestra particular podrían ser ligeramente diferentes a las de otra muestra. Una tercera muestra sería también diferente, como lo serían también una cuarta, una quinta, etc. Las muestras estadísticas descriptivas sólo estiman los valores de la población, por lo que las deducciones realizadas deben tener en cuenta qué posibles muestras estadísticas pueden haber sido desarrolladas.

Veamos el siguiente ejemplo (figura 9.1). Imaginemos que un investigador está interesado en determinar el nivel de comprensión lectora de alumnos de noveno curso de una escuela. Se selecciona una muestra aleatoria; imaginemos que la media de la muestra es 65, con una desviación típica de 2.5 (figura 9.1a). Ahora imaginemos que el investigador selecciona al azar otra muestra y esta vez la media es 66. Ahora existen dos medias (figura 9.1b). ¿Cuál es la correcta? Decide seleccionar cinco muestras más y consigue medias de 64, 63.5, 64.5, 65.7, y 65.2 (figura 9.1c).

FIGURA 9.1: MUESTRA ALEATORIA CON (a) UNA MEDIA (b) DOS MEDIAS (c) VARIOS MEDIAS Y (d) LOS MEDIAS MOSTRANDO UNA CURVA NORMAL



Ahora bien, ¿cuál es la correcta? El investigador decide averiguarlo realmente y toma treinta muestras más. ¡Seguro que lo hará! Se obtienen diferentes muestras, pero cuando se unen las medias empiezan a parecerse a algo que resulta familiar: una curva normal (figura 9.1d).

De hecho, si se extraen 100 muestras, cuando se unen sus medias formarían una curva normal, con su propia media y desviación típica. La distribución resultante se denomina la **distribución de la muestra** y a la desviación típica se le da el nombre de **error típico**. De manera que, cuando imaginamos que se extraen y se puntúan diferentes muestras de una población, el error típico es la desviación típica de esta distribución muestral. Pueden, pues, calcularse la *media de medias* y el error típico de la distribución muestral y el investigador puede usar esta información para saber, no cuál es la media de la población, sino su probabilidad teniendo un cierto valor basado en las propiedades de la curva normal. De modo que cerca de dos tercios de las medias obtenidas estarían dentro de un error típico de la media de medias y el 96 por ciento de las medias estarían dentro de dos desviaciones típicas. Por lo tanto, ahora el investigador puede describir un rango de medias probables y estar completamente seguro de éste, aunque las medias de muestras particulares tengan ligeras diferencias entre sí.

En realidad, un investigador no suele tomar más de una muestra, pero se puede calcular la media de la población a partir de la media y la varianza de la muestra. Por lo tanto, es aquí donde se integra la probabilidad y la Estadística inferencial. La Estadística inferencial se emplea para estimar la probabilidad de que la media de una población se encuentre dentro de una serie de valores posibles. El investigador es capaz de inferir que la media de la población estará 68

veces de cada 100 dentro de un error típico de la media de medias y que estará dentro de dos errores típicos 96 veces de cada 100.

Un concepto relacionado que se está utilizando cada vez más es *intervalo de confianza*. El **intervalo de confianza** (IC) es una serie de valores numéricos en los que probablemente reside el valor real de la población. Los límites superior e inferior del intervalo de confianza se llaman *límites confidenciales*. Normalmente, los investigadores informan de un intervalo de confianza de 95 o un 99 por ciento, lo que significa que la probabilidad de que la media de la población se encuentre entre los límites confidenciales es de un 95 o 99 por ciento, respectivamente. Por ejemplo, si se dice que el intervalo de confianza de 99% de una encuesta estatal es .435-.492, esto significa que a partir de la muestra se puede inferir que hay un 99 por ciento de probabilidad de que entre un 43.5 y un 49.2 por ciento de la media de la población esté a favor de una propuesta determinada. De manera que es muy probable que la mayoría de la población esté en contra de la propuesta. Si el intervalo de confianza al 99 por ciento estuviera completamente por encima de .50, entonces una mayoría estaría probablemente a favor.

Supongamos que un investigador no está interesado en extraer una muestra de una población, sino que incluye a toda la población como muestra. ¿Significaría esto que el investigador podría ignorar los principios del error muestral y la Estadística inferencial al obtener un resultado que representa al grupo entero? Aunque en este caso el error muestral no sería un problema, el error de medida sí que lo es. Recordemos cuando se decía en el capítulo 6 que cada vez que establecemos variables en la educación, sus mediciones no son nunca perfectas. Siempre hay algún grado de error, resumido estadísticamente como el error típico de medida. De manera que deducimos un valor real o auténtico de una variable a partir de una medida imperfecta. Esto podría considerarse como un tipo de error de muestreo en el sentido de que la medida obtenida es una «muestra» que estima el valor real. Por consiguiente en toda investigación educativa se da definitivamente el error de medida y en algunas investigaciones se da también el error de muestreo.

Esta sección del capítulo ha explicado la lógica de la Estadística inferencial mediante la estimación de las medias de la población a partir de las medias de la muestra y ha presentado el concepto de intervalos de confianza. Sin embargo, en la mayoría de las investigaciones, estamos interesados en mucho más que en estimar poblaciones a partir de muestras. Normalmente, queremos comparar medias poblacionales entre ellas o con algunos valores establecidos. La próxima sección trata de cómo se hacen estas comparaciones.

HIPÓTESIS NULA

Imaginemos que un investigador quiere comparar las actitudes de alumnos de sexto curso hacia la escuela con las de alumnos de cuarto curso. El investigador

selecciona al azar muestras de alumnos de sexto y de cuarto curso y calcula las medias de cada grupo. Éstas son 30 para los alumnos de cuarto y 37 para los de sexto. ¿Puede, por lo tanto, el investigador llegar a la conclusión de que los alumnos de sexto curso tienen actitudes más positivas que los de cuarto curso?

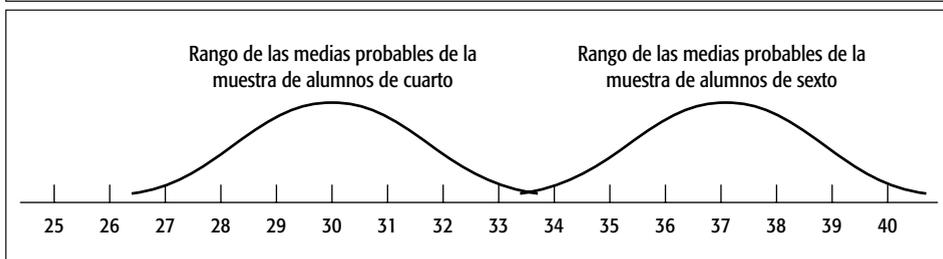
Quizá, pero esta conclusión debe tener en cuenta el error de muestreo y de medida. De este modo, se calculan y comparan las medias de la población para encontrar la probabilidad de que las posibles medias de la población de cada grupo sean diferentes. Las probabilidades se formalizan mediante enunciados que son sometidos a examen. Se hace referencia a estos enunciados como *hipótesis*. Ya hemos presentado las hipótesis de investigación como una predicción de la investigación que es puesta a prueba (en este ejemplo, la hipótesis de investigación podría ser que los alumnos de sexto curso tienen actitudes más positivas que los alumnos de cuarto curso).

Cuando nos referimos a la probabilidad en términos de error de muestreo y medida, el enunciado empleado se denomina **hipótesis estadística**.

Con frecuencia, las hipótesis estadísticas se enuncian en lo que se denomina forma nula. Las **hipótesis nulas**, normalmente implícitas en las hipótesis de investigación en vez de expresadas explícitamente, enuncian que no existen diferencias entre las medias de la población de los dos grupos. Es decir, las medias poblacionales son las mismas. El investigador emplea una prueba estadística inferencial para determinar la probabilidad de que la hipótesis nula no sea cierta. Si es falsa, entonces hay una alta probabilidad de que exista una diferencia entre los dos grupos. La hipótesis nula en nuestro ejemplo sería que las actitudes de los alumnos de sexto y cuarto curso hacia la escuela son las mismas. Si podemos demostrar que hay una alta probabilidad de estar en lo cierto al rechazar la hipótesis nula, entonces tendremos evidencia de existen diferencias en las actitudes de los alumnos.

En teoría, sabemos que puede estimarse el rango de medias de ambos grupos de la población y que, si hay un pequeño solapamiento en estos rangos, entonces es probable que las medias de la población sean diferentes. Este caso se refleja en la figura 9.2. Virtualmente no hay solapamiento entre las dos curvas normales

FIGURA 9.2: RANGO DE LAS MEDIAS DE LA POBLACIÓN DE LOS DOS GRUPOS



Esto significa que podemos estar seguros de no equivocarnos al rechazar la hipótesis nula.

La razón por la que la hipótesis nula se utiliza en Estadística inferencial es que nunca probamos que algo sea cierto, sino que más bien no conseguimos probar que es falso. El no conseguir demostrar su falsedad coincide con la realidad de la probabilidad en nuestras vidas. En otras palabras, si no podemos encontrar una evidencia convincente de que son diferentes, la conclusión más verosímil es que son iguales. La hipótesis de investigación es mucho más importante que la hipótesis nula en la creación y el diseño de la investigación. La hipótesis nula es una necesidad técnica en el uso de la Estadística inferencial.

También habría que precisar que el no lograr rechazar la hipótesis nula no tiene que significar necesariamente que sea cierta. Es especialmente difícil aceptar hipótesis nulas como ciertas en estudios que utilizan un escaso número de sujetos o pruebas con una baja fiabilidad. El hecho de que no se rechace la hipótesis nula puede ser consecuencia de un gran error de muestreo o error de medida, o bien, puede atribuirse a otros muchos factores que están fuera del alcance de este libro. En pocas palabras, es conveniente ser cautos con respecto a las conclusiones que se basan en el rechazo de la hipótesis nula.

NIVEL DE SIGNIFICACIÓN

Como la base de la Estadística inferencial es la probabilidad de una estimación, entonces aceptar o rechazar la hipótesis nula está también relacionado con la probabilidad, en lugar de ser una decisión dicotómica. Es decir, debido a los errores de muestreo o de medida, podemos dar sólo la probabilidad de tomar o no la decisión correcta al acertar o al rechazar la hipótesis nula. En otras palabras, podemos estar completamente seguros de que, de cada cien, un cierto número de veces las medias muestrales que podríamos obtener no serán las mismas.

El **nivel de significación** se utiliza para indicar cuál es la probabilidad de que estemos equivocados al rechazar la hipótesis nula. También llamado nivel de probabilidad, o nivel p , se representa con un decimal y nos dice cuántas veces de cien o de mil nos equivocaríamos al rechazar la hipótesis nula suponiendo que ésta fuera cierta (en otras palabras, cuántas veces no esperaríamos ninguna diferencia real aunque rechazáramos la hipótesis nula). La lógica del nivel de significación es que suponemos que la hipótesis nula es cierta para, a continuación, examinar cuál es la probabilidad de que la media de la muestra que hemos calculado sea diferente sólo por casualidad. Si observamos que existe la probabilidad de una sola vez entre cien de que encontremos una diferencia particular en las medias por casualidad o por fluctuaciones aleatorias, ($p = .01$), entonces, probablemente rechazaríamos la hipótesis nula porque es bastante probable que sea

falsa. En otras palabras, el nivel de significación nos indica la probabilidad ocasional de encontrar diferencias entre las medias. Por lo tanto, cuanto más bajo es el nivel de significación, más confianza tendremos en que estaremos seguros rechazando la hipótesis nula. Después de todo, por ejemplo, si encontramos una diferencia de cinco puntos entre dos medias que, a través de nuestra hipótesis nula, suponemos que son iguales y nuestras estadísticas nos dicen que hay sólo una posibilidad entre cien de encontrar una diferencia de cinco puntos de manera casual ($p = .001$), entonces, lo más lógico es suponer que la hipótesis nula es falsa y rechazarla (¡o decir que somos muy, muy afortunados!). Rechazamos la hipótesis nula a favor de la hipótesis de investigación, también denominada hipótesis alternativa.

Algunos investigadores pueden emplear un enfoque para la comprobación de hipótesis que tiene ligeras diferencias: antes de la recogida de datos se establece un nivel de significación determinado para que actúe como un criterio para la aceptación o no de la hipótesis nula. Este «valor crítico», se llama el nivel alfa (α) y determina la conclusión basada en los números obtenidos por los resultados. Sin embargo, usar sólo el nivel alfa da lugar a una decisión dicotómica, mientras que la comprobación de la significación con los valores p aporta más información.

Errores en la comprobación de la hipótesis

El propósito de la Estadística inferencial, la hipótesis nula y los niveles de significación es tomar una decisión, basada en la probabilidad, sobre la naturaleza de las poblaciones y los valores reales de las variables. Es posible que la decisión sea equivocada. Cuando la decisión es rechazar la hipótesis nula cuando de hecho la hipótesis nula es cierta, el investigador ha cometido lo que se denomina un **error tipo I**. La probabilidad de cometer este tipo de error equivale al nivel de significación: es decir, con un nivel de significación del .05 hay una probabilidad de cinco veces de cada cien de que los datos de la muestra conduzcan al investigador a rechazar la hipótesis nula cuando de hecho es cierta. Por consiguiente, un investigador evita el error tipo I en el grado en el que el nivel de significación es alto (es decir, un nivel .001 es mejor que .01 para evitar los errores tipo I).

Se comete otro tipo de decisión equivocada cuando no se rechaza la hipótesis nula, siendo ésta realmente falsa. A esto se le llama **error tipo II**. Aunque no hay ninguna relación directa entre el nivel de significación y la probabilidad de cometer un error tipo II, cuando aumenta el nivel de significación la probabilidad del error tipo II disminuye. De manera que un nivel de significación de .10 es mejor para evitar un error tipo II que .05 o .01. La figura 9.3 muestra cómo el tipo de error está relacionado con los tipos de decisiones.

Interpretación del nivel de significación

La interpretación de rechazar o no una hipótesis nula depende de si el investigador está interesado en evitar un error tipo I o tipo II, y en si ha establecido previamente un nivel alfa.

FIGURA 9.3: RELACIÓN DEL ESTADO DE LA NATURALEZA, LAS DECISIONES Y EL ERROR EN LA COMPROBACIÓN DE HIPÓTESIS

	Estado de su naturaleza	
	La hipótesis nula es cierta	La hipótesis nula es falsa
Rechazo de la hipótesis nula	Error tipo I	Decisión correcta
No se consigue rechazar la hipótesis nula	Decisión correcta	Error tipo II

Si se fija un valor predeterminado, como .05 o .01 para un error de tipo I, entonces el investigador rechaza la hipótesis nula mediante la comparación del nivel de significación calculado con el nivel predeterminado. Si la significación calculada es menor que el nivel predeterminado (por ejemplo, $.01 < .05$) entonces se rechaza la hipótesis nula.

En muchos estudios de investigación no hay un nivel alfa predeterminado. En estos estudios, los estadísticos utilizan una regla general para rechazar la hipótesis nula. Si el valor p es igual o menor que .05, entonces se rechaza la hipótesis nula y se llega a la conclusión de que hay una diferencia **estadísticamente significativa** (aunque para ser más precisos siempre hay alguna diferencia en algún nivel de confianza). Un valor p entre .05 y .10 se considera, con frecuencia, como *marginalmente* significativo, y algo mayor de .10 se considera una diferencia no significativa. Por lo tanto, lo que estamos diciendo es que si hay más de una posibilidad entre diez de estar equivocados al rechazar la hipótesis nula (una posibilidad entre diez de que las medias sean iguales) entonces es demasiado arriesgado decir que las medias son diferentes. Los resultados pueden deberse más a un error que al efecto de un tratamiento o a una diferencia real.

Es mejor informar del nivel p para cada prueba estadística porque las convenciones para rechazar la hipótesis nula son «reglas del pulgar» generales. Los investigadores individuales y los consumidores, según las circunstancias, pueden discrepar con respecto a lo que constituye una diferencia estadísticamente significativa. Por ejemplo, un nivel de .05, en general aceptado como estadísticamente significativo, sería probablemente inaceptable si la prueba implicara el uso de una droga que podría causar la muerte (el investigador está equivocado cinco veces de cada cien al decir que no ocurrirá ninguna muerte). Ante esta situación, un nivel de .000001 podría considerarse, de forma más apropiada, estadísticamente significativo.

Otro aspecto importante en la interpretación de los niveles p y las conclusiones correspondientes es que, aunque para los investigadores es común no conseguir rechazar la hipótesis nula (ejemplo $p = .20$), el hecho de no encontrar una diferencia o relación estadísticamente significativa no tiene que representar necesariamente que «en realidad» no existe ninguna diferencia ni relación. Sólo cuando las premisas que justifican la investigación (lo que se denomina «potencia» de la prueba) no son consideradas datos significativos, evidencian que no hay ninguna relación. La razón por la que un resultado no significativo es, a menudo, ininterpretable es que muchos factores, como la baja fiabilidad, la difusión del tratamiento, un número insuficiente de sujetos, etc., pueden causar la falta de significación. De este modo, tendemos a creer que un resultado significativo indica que existe una relación real, pero lo contrario no es necesariamente cierto. Especialmente en casos con un tamaño muestral pequeño (lo que dificulta encontrar diferencias significativas), un resultado no significativo debería interpretarse como indicador de que es necesario proseguir la investigación, antes que como falta de relación.

Una de las cuestiones más importantes que crean confusión en la interpretación estadística es la decisión de si los resultados son representativos. Es decir, ¿hasta qué punto los resultados son diferentes en el mundo real? ¿Son significativos los resultados desde el punto de vista educativo y no sólo desde el punto de vista estadístico? Las pruebas estadísticas nos dicen sólo que existe una diferencia, pero debe también considerarse el valor o la importancia de un resultado. Cuando se afirma que un dato es estadísticamente significativo, el lector debería examinar las medias descritas para tener en cuenta cómo son de diferentes y la magnitud de las correlaciones. La representatividad se relaciona con lo específico de una situación. Por ejemplo, puede haber una diferencia, estadísticamente significativa, en el rendimiento lector entre alumnos de primer curso que utilizan el currículum X frente a los que utilizan el currículum Y, pero eso no significa que debiera adoptarse el currículum X. Es posible que la diferencia represente sólo un percentil y que el currículum X cueste muchos miles de dólares más que el currículum Y. Sólo el lector puede juzgar lo que es representativo. Muchos in-

investigadores tienden a asumir de manera automática que significación estadística equivale a significación educativa, pero esto es simplemente falso.

Una de las razones de que valores p estadísticamente significativos puedan ser representativos es que el valor calculado está directamente relacionado con el tamaño muestral. Por lo tanto, es posible tener una muestra muy amplia, una diferencia o relación muy pequeña y, sin embargo, considerarlo como «significativo». Por ejemplo, una correlación de .44 será estadísticamente significativa en el nivel .05 con un tamaño muestral tan pequeño como 20, mientras que una muestra de 5.000 permitirá un resultado estadísticamente significativo de .05 con una correlación de sólo .028, que no es, en la práctica, relación alguna.

La *American Psychological Association* (1999) y varias revistas científicas recomiendan ahora insistentemente o requieren que los investigadores presenten indicadores apropiados que ilustren la fuerza o la magnitud de una diferencia o relación junto con medidas de significación estadística. Estas «medidas de la magnitud del efecto», como son llamadas, son medidas de fuerza de la asociación o tamaño del efecto. Las medidas de la asociación se utilizan para estimar proporciones de varianza común, parecidas al coeficiente de determinación. El tamaño del efecto se utiliza con más frecuencia. Habitualmente se presenta de forma general como la ratio de la diferencia entre las medias de grupo dividida entre la desviación típica estimada de la población. Según Cohen (1988), entonces el *índice del tamaño del efecto* aporta una indicación de la diferencia práctica o representativa. Los índices del tamaño del efecto cercanos a .20 se consideran pequeños; los que están sobre .5 como efectos moderados o medianos, y los de .8 y más como efectos grandes. En el extracto 9.1 aparece un ejemplo de tamaño del efecto descrito en una revista. Por supuesto, las medidas de magnitud del efecto, junto con los intervalos de confianza, aportan una información mucho mejor que un simple rechazo de la hipótesis nula.

El extracto 9.2 expone niveles de significación para mostrar cuál de los resultados es estadísticamente significativo. Recuerde que pueden usarse varias convenciones para informar del nivel de significación, como niveles de significación, nivel p y alfa (?). Algunos investigadores pueden exponer que la probabilidad (p) es menor que ($<$) un valor específico (por ejemplo, $p < .05$ o $p = .001$) en vez de presentar el nivel real de significación.

COMPARACIÓN DE DOS MEDIAS: LA PRUEBA t

Hay muchas situaciones en la investigación en las que se compara la media de un grupo con la media de otro para determinar la probabilidad de que las medias de las poblaciones correspondientes son diferentes.

EXTRACTO 9.1 PRESENTACIÓN DEL TAMAÑO DEL EFECTO

Según las categorías de Cohen de tamaños de efecto pequeño, mediano y grande (Buzz, 1995), un análisis de potencia reveló que la mayoría de los tamaños del efecto para las diferencias de género en este estudio iban de pequeño a mediano (ver tabla 3).

TABLA 3: Resumen de datos de igualdad de sexos

Dependiente Medida	Hombre		Mujer		F	gl	p	Tamaño de efecto	Categoría Cohen
	M	Dt	M	Dt					
Profesor llamando a los alumnos	0.2480	0.2425	0.2034	0.2608	0.71	1.22	.4131	0.18	Pequeño
Estudiante voluntario	1.1000	1.0615	0.9548	1.3027	0.45	1.22	.5081	0.12	Pequeño
Profesor interactuando con los alumnos	1.3274	1.2999	1.5623	1.6247	0.55	1.22	.5078	0.11	Pequeño
Alumnos levantando la mano	0.1835	0.1469	0.1334	0.1155	1.70	1.22	.2055	0.38	Mediano
Quedándose después de clase	0.1440	0.0534	0.1066	0.0612	3.18	1.22	.0885	0.58	Mediano
Lugar de asiento	1.2496	1.4105	0.3891	0.0571	1.65	1.28	.2101	1.17	Grande

Nota: Los valores que aquí aparecen están en forma de ratio. Las medias de las alumnas se basan en el promedio de la respuesta de una niña entre el número de niñas en la clase. Las medias de los alumnos varones se basan en los resultados de la respuesta de los varones dividida entre el número de niños en clase. Niños y niñas son comparados basándose en sus porcentajes de participación real, teniendo en cuenta la proporción de niñas y niños en clase.

Fuente: Brady, K. L. y Eisler, R. M. (1999), «Sex and gender in the college classroom: A quantitative análisis of faculty-study interactions and perceptions», *Journal of Educational Psychology*, 91 (1), 127-145.

EXTRACTO 9.2 NIVELES DE SIGNIFICACIÓN

En las tablas 3 y 4 se presentan las medias ajustadas y las desviaciones típicas para los recuerdos y las medidas de exactitud. El análisis de los resultados del recuerdo indi-

caron que a lo largo de los pasajes (inicial y transferencia), hay un efecto principal significativo del entrenamiento para hacer tests $F(1,68) = 5.32, p = .02$. Los grupos que recibieron el entrenamiento cooperativo tuvieron una tasa de recuerdo más alta que aquéllos que recibieron el entrenamiento individual. El efecto principal del entrenamiento en estudio cooperativo frente al individual alcanzó significación, $F(1, 68) = 3.31, p = .07$. No hubo interacciones significativas.

El análisis de los resultados de la exactitud revelaron que a lo largo de los pasajes, el efecto principal del entrenamiento en estudio cooperativo frente al individual fue significativo, $F(1, 68) = 4.70, p = .03$. La interacción entrenamiento por pasaje resultó también significativa, $F(1, 70) = 5.89, p = .01$. Comparaciones posteriores demostraron que los grupos que estudiaron cooperativamente el pasaje 1 recordaban de manera significativa más información exacta que los grupos que estudiaron el pasaje 1 de manera individual ($p < .05$). Ninguna otra diferencia entre las medias de los grupos resultó significativa.

Fuente: Lambiotte, J. G. y otros (1987), «Cooperative Learning and test taking: Transfer of skills», *Contemporary Educational Psychology*, 12 (1).

El procedimiento estadístico más común para determinar el nivel de significación al comparar dos medias es la **prueba t** . La prueba t es una fórmula que genera un número que se utiliza para determinar el nivel de probabilidad (nivel p) de rechazar la hipótesis nula.

Lo que ocurre es que las medias de las muestras, las desviaciones típicas y los tamaños muestrales se utilizan en la ecuación de la prueba t para obtener un valor t (llamado a veces estadístico t). La fórmula para calcular el valor t es:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

Por lo que

\bar{X}_1 = media del grupo 1

\bar{X}_2 = media del grupo 2

$S_{\bar{X}_1 - \bar{X}_2}$ = error típico de la diferencia entre las medias

El error típico de la diferencia entre las medias puede ser concebido como una medida de la cantidad de error al estimar la media de la población a partir de la media muestral. Conforme aumenta la distancia entre \bar{X}_1 y \bar{X}_2 , como el error implicado en la estimación de las medias disminuye, el estadístico t es mayor. El valor calculado t es un número de tres o cuatro dígitos con dos decimales, como 2.30; 3.16; 8.72; 1.85. Para determinar el nivel de significación, el investigador compara este número con los valores t teóricos en una tabla. La tabla se denomina distribución t , o valores críticos para la prueba t , y se encuentra en la tabla 9.1. El investigador utiliza la tabla localizando dos números: los **grados de libertad** (gl) y

TABLA 9.1: DISTRIBUCIÓN t

	Nivel de significación para un contraste de una cola				
	.05	.025	.01	.005	.0005
gl	Nivel de significación para un contraste de dos colas				
	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.619
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.408
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

Fuente: Fisher y Yates (1974), *Statistical Tables for Biological, Agricultural and Medical Research*, tabla III, 6ª edición, Londres: Longman Group UK Ltd. (publicado previamente por Oliver and Boyd Ltd. Edimburgo). Reimpreso con permiso de los autores y editores.

el nivel de significación deseado. El término *grados de libertad* es un concepto matemático que denota el número de observaciones independientes que están libres de variaciones. Para cada prueba estadística hay un número correspondiente de grados de libertad a calcular para, a continuación, utilizar este número para calcular la significación estadística de la prueba. En la distribución de una tabla t , el número en la intersección de la fila de los grados de libertad y la columna del nivel de significación es el valor teórico de t . Si esta t crítica es menor que el valor t calculado por la ecuación de la prueba t , esto significa que la diferencia observada en las medias es mayor de lo que podía haberse esperado bajo la hipótesis nula, de modo que la hipótesis puede ser rechazada en ese nivel de significación.

Observe que en la parte superior de la tabla hay dos filas, la de arriba para lo que se conoce como «test de significación de una cola o unilateral» y la otra para el «test de dos colas o bilateral». Las «colas» hacen referencia a los extremos de la distribución muestral normal y se utilizan en el examen de la significación. «Una cola» significa que uno de los dos extremos de la distribución se utiliza como región de rechazo, de manera que si el valor p es .05, este 5 por ciento se encuentra en un extremo de la distribución. En una prueba de dos colas la región se divide entre los dos lados de la distribución, por ejemplo, para un valor p de .05, el .025 por ciento está en cada uno de los extremos de la distribución. El contraste de una cola es más liberal y debería utilizarse sólo cuando el investigador confía en gran medida en que no obtendrá un resultado que sea contrario a la hipótesis de investigación. A menos que se diga lo contrario, puede considerarse generalmente que las pruebas de significación son bilaterales.

El estadístico t calculado y el nivel p correspondiente se exponen en la mayoría de los estudios. En el extracto 9.3 se reproduce la sección de resultados de una investigación para mostrar cómo se resumen las pruebas t . El número entre paréntesis que sigue a la t son los grados de libertad.

Existen dos formas diferentes de la ecuación utilizada en la prueba t , una para muestras independientes y otra para muestras no independientes. Las muestras independientes son grupos de sujetos que no tienen ninguna relación entre ellos; las dos muestras tienen sujetos diferentes en cada grupo y, normalmente, los sujetos son designados de manera aleatoria de una población semejante o tomados de dos poblaciones diferentes. Por lo tanto, si un investigador está comprobando la diferencia entre las medias de un grupo experimental y un grupo de control en un diseño de sólo postest, la prueba **t para muestras independientes** sería la apropiada. Al comparar las actitudes de alumnos de cuarto y sexto curso, se utilizaría también una prueba t para muestras independientes.

Se puede hacer referencia a la segunda forma de la prueba t mediante diferentes nombres, incluyendo prueba t emparejada, de muestras dependientes o de muestras relacionadas. Esta prueba t se utiliza en situaciones en las que los sujetos de los dos grupos están emparejados o equiparados de algún modo. Un ejemplo

EXTRACTO 9.3**PRUEBA 7 PARA MUESTRAS INDEPENDIENTES**

Medias, desviaciones típicas y estadísticos t obtenidos para comparar la evaluación entre la educación a distancia y el curso en el campus aparecen representadas en la tabla 3.

TABLA 3: Medias, desviaciones típicas y estadísticas t para comparar la evaluación entre la educación a distancia y el curso en el campus

Área de evaluación	SPED 6126			SPED 6127		
	Campus (n = 4)	Distancia (n = 23)	t obtenida	Campus (n = 11)	Distancia (n = 13)	t obtenida
Puntuación general						
M	3.69	3.94	-0.81	3.69	3.79	-0.48
Dt	0.59	0.33		0.28	0.44	
Curso						
M	3.56	3.88	-1.82	3.72	3.60	1.14
Dt	0.33	0.31		0.29	0.22	
Profesor						
M	3.65	3.88	-0.75	3.65	3.65	0.00
Dt	0.59	0.34		0.19	0.43	
Organización						
M	4.15	4.23	-0.44	3.83	4.25	-6.70*
Dt	0.34	0.17		0.19	0.10	
Enseñanza						
M	3.48	3.62	-0.56	3.58	3.42	1.09
Dt	0.47	0.44		0.44	0.23	
Comunicación						
M	3.49	3.79	-1.26	3.56	3.63	-0.54
Dt	0.48	0.27		0.13	0.40	

* $p < .01$.

Fuente: Spooner, F., Jordan, L., Algozzine, B. y Spooner, M. (1999), «Student ratings of instruction in distance learning and on-campus courses», *Journal of Educational Research*, 92 (3), 132-140.

frecuente de este caso es el mismo grupo de sujetos sometido a examen dos veces, como en un estudio de pretest-postest. Ya estén en cada grupo los mismos sujetos o sujetos diferentes, mientras haya una relación sistemática entre los grupos es necesario usar la prueba t de **muestras dependientes** para calcular la probabilidad de rechazo de la hipótesis nula.

Aunque los grados de libertad y las fórmulas son diferentes para cada una de las formas de la prueba t , la interpretación y la presentación de los resultados es la misma (los «gl» para la prueba t dependiente es el número de parejas menos uno). Por lo tanto, el lector de investigación no tiene que preocuparse acerca de si se ha utilizado la fórmula correcta.

La prueba t puede utilizarse para propósitos diferentes a la comparación de las medias de dos muestras. La prueba t se emplea cuando un investigador quiere demostrar que un coeficiente de correlación es significativamente diferente de 0 (ninguna correlación). Se puede comparar la media de un grupo con un número en vez de con otra media, siendo posible también comparar varianzas en lugar de medias. Como hay tantos usos para la prueba t la encontramos con frecuencia en investigación.

Una explicación más concreta del uso de la prueba t es el ejemplo siguiente. Supongamos que un investigador está interesado en averiguar si existe una diferencia significativa entre los alumnos de sexto curso con ojos azules y con ojos marrones con respecto a su comprensión lectora. La pregunta de investigación sería: ¿hay diferencia en la comprensión lectora (la variable dependiente) de los alumnos de cuarto con ojos azules en comparación con los alumnos de cuarto curso con ojos marrones (la variable independiente)? La hipótesis nula sería no hay ninguna diferencia entre los alumnos de cuarto curso con ojos azules y ojos marrones en su comprensión lectora. Para poner a prueba esta hipótesis el investigador seleccionaría aleatoriamente una muestra de alumnos de cuarto curso con ojos azules y con ojos marrones de la población de todos los alumnos de cuarto curso. Supongamos que la media de comprensión lectora de la muestra de alumnos con ojos azules es 54 y la media de la muestra de alumnos de cuarto con ojos marrones es 48. Como damos por supuesta la hipótesis nula –que las medias de la población son iguales– empleamos la prueba t para mostrar con cuánta frecuencia tendría lugar la diferencia de puntuación en las muestras si las medias de la población fueran iguales. Si nuestros grados de libertad (el tamaño muestral total menos 1) son 60 y el valor t calculado 2.00, podemos ver remitiéndonos a la tabla 9.1 que la probabilidad de alcanzar esta diferencia en las medias de la muestras en un contraste bilateral es .05, o cinco veces de cada cien. Rechazamos la hipótesis nula y decimos que existe una diferencia estadísticamente significativa entre la comprensión lectora de los alumnos de cuarto curso con ojos azules y marrones.

COMPARACIÓN DE DOS O MÁS MEDIAS: ANÁLISIS DE VARIANZA

Análisis de varianza de un factor o una vía

Si se realiza un estudio en el que se comparan dos o más medias muestrales en una variable independiente, para poner a prueba la hipótesis nula el investigador

emplearía un procedimiento denominado análisis de varianza de un factor (su abreviatura es ANOVA). El ANOVA es simplemente una extensión de la prueba t . En lugar de que el investigador emplee múltiples pruebas t para comparar todas las posibles parejas de medias en un estudio de dos o más grupos, el ANOVA le permite comprobar las diferencias entre todos los grupos y llegar a conclusiones de probabilidad más precisas que cuando se utilizan una serie de pruebas t por separado. Esto se denomina **análisis de varianza** porque la fórmula estadística utiliza las varianzas de los grupos y no sus medias para calcular un valor que refleje el grado de diferencias entre sus medias. En vez de un estadístico t , el ANOVA calcula un estadístico F (o puntuación F). La F es análoga a la t . Es un número de tres o cuatro dígitos que se utiliza en una distribución de una tabla F con los grados de libertad para encontrar el nivel de significación que el investigador utiliza para rechazar o no la hipótesis nula. Existen dos grados de libertad. El primero es el número de grupos en el estudio menos uno y el segundo es el número total de sujetos menos el número de grupos. Estos números siguen la F en la presentación de los resultados de un ANOVA. Por ejemplo, al presentar $F(4,80) = 4.25$, los grados de libertad significan que se comparan las medias de cinco grupos y 85 sujetos en el análisis.

El ANOVA plantea la siguiente pregunta: ¿existen diferencias significativas entre las medias de la población? Si el valor F calculado es lo suficientemente grande, entonces la hipótesis nula (que significa que no hay ninguna diferencia entre los grupos) puede rechazarse confiando en que el investigador no se equivoca al concluir que, al menos, dos medias son diferentes. Supongamos, por ejemplo, que un investigador está comparando el locus de control de tres grupos: alumnos de alto, medio y bajo rendimiento. El investigador selecciona una muestra aleatoria de cada grupo, administra pruebas para medir el locus de control y calcula las medias y varianzas de cada grupo. Vayamos aun más lejos suponiendo que las medias de los grupos son A (bajo rendimiento) = 18; B (rendimiento medio) = 20, y C (alto rendimiento) = 25. Por lo tanto, la hipótesis nula puesta a prueba es que las medias de la población de 18, 20, y 25 son iguales o, mejor dicho, que son diferentes solamente a causa de errores de muestreo y medida. Si la F fuera 5.12 y $p < .01$, entonces el investigador puede llegar a la conclusión de que al menos dos de las medias son diferentes y de que esta conclusión será correcta 99 veces de cada 100.

Generalmente, los resultados de un ANOVA de un factor se presentan indicando, en la sección de resultados, los grupos que son diferentes. Una tabla de medias y desviaciones típicas acompañará a la información escrita. En la presentación de los resultados, el investigador en ocasiones escribirá dos números delante del ANOVA. Este será un primer número, un signo de multiplicación y a continuación otro dígito (por ejemplo, 1×3 ; 1×5 ; 1×2). Esto significa que hay una variable independiente (1) que tiene el número de grupos o niveles indicado mediante el segundo número. Por lo tanto, un ANOVA 1×4 es un ANOVA de un factor que compara las medias de cuatro grupos.

Procedimientos de comparaciones planificadas y *post hoc*

Cuando un investigador utiliza un ANOVA para comprobar la hipótesis nula de que tres medias son iguales, el valor F que resulta estadísticamente significativo indica al investigador sólo que dos o más de las medias son diferentes. Con frecuencia, el investigador necesita utilizar exámenes estadísticos más detallados que le indicarán cuáles son las medias que difieren entre sí. Estas pruebas se denominan **comparaciones *post hoc***. Otros términos sinónimos son pruebas *a posteriori* y comparaciones múltiples. Otro procedimiento, denominado *la técnica Bonferroni* se utiliza también como una prueba *post hoc*, aunque habitualmente se utiliza cuando los investigadores están estudiando muchas variables dependientes para ajustar el nivel de significación y reducir la posibilidad de encontrar una diferencia significativa a causa de múltiples pruebas estadísticas. Están diseñadas para comprobar cada posible pareja de medias.

Hay cinco pruebas de comparaciones múltiples de uso común: LSD de Fisher (acrónimo de *Least Significant Difference*, que puede ser traducido como test de la diferencia mínima significativa), el test de Duncan, también denominado test de rango múltiple, el test de Newman-Keuls, HSD de Tukey (cuyas iniciales corresponden a *Honestly Significant Difference*, traducido como test de la diferencia francamente significativa) y el test de Scheffé. Todas las pruebas se utilizan de la misma manera, pero difieren en la facilidad con la que se obtiene una diferencia significativa; es decir, que en algunas pruebas las medias deben estar más alejadas entre sí que en otras para que la diferencia sea estadísticamente significativa. Las pruebas que requieren una mayor diferencia entre las medias se consideran conservadoras, en tanto que aquéllas que permiten una diferencia menor se denominan liberales. La lista de las pruebas anteriores es, en este sentido, secuencial, con el test de Fischer considerado como el más liberal y el de Scheffé el más conservador. Las dos pruebas más comunes son la de Tukey y la de Scheffé, pero se pueden extraer diferentes conclusiones de un estudio dependiendo de la técnica de comparaciones múltiples empleada.

El lector puede ahora preguntarse por qué razón un investigador no utiliza diferentes pruebas t tras el hallazgo de una F significativa. La respuesta es que si utilizara múltiples pruebas t , el investigador aumentaría la probabilidad de encontrar por casualidad una diferencia significativa donde no existe ninguna. Por ejemplo, si la investigación implica calcular 100 pruebas t y el nivel de significación es 0.05, el investigador puede esperar que 5 de cada 10 pruebas t sean significativas de manera casual. Las pruebas *a posteriori* controlan este problema teniendo en cuenta el número de comparaciones realizadas. En algunos estudios no se emplean pruebas *a posteriori* a pesar de que la F sea significativa. Esto sucede porque, evidentemente, el mero examen visual de las medias de los grupos muestra cuáles son diferentes. Por ejemplo, si las medias de tres grupos son 15, 20 y 20.5,

y la puntuación F es significativa con un nivel p del .01, está bastante claro dónde residen las diferencias y no hay ninguna necesidad de pruebas *a posteriori*. Debemos destacar un aspecto: es imposible interpretar una prueba t o análisis de varianza sin examinar las medias de los grupos; para que la Estadística inferencial tenga un significado es esencial examinar las medias muestrales.

El extracto 9.4 es un ejemplo del uso de un ANOVA de una vía y de pruebas *a posteriori*.

EXTRACTO 9.4 **ANOVA Y PRUEBAS A POSTERIORI**

Se empleó el análisis de varianza de una vía para analizar los datos de rendimiento. Allí donde se encontraron diferencias significativas, se utilizaron pruebas SNK para determinar qué medias diferían significativamente.

Nuestra segunda hipótesis establecía que los grupos enseñados con el libro infantil obtendrían resultados significativamente más altos en el postest que el grupo enseñado únicamente con el libro de texto. Los datos respaldaban la hipótesis, $F(2,54) = 10.23$, $p = .0002$. La media y la desviación típica para el grupo del libro infantil eran 8.79 y 4.66, respectivamente. La media y desviación típica para el grupo de libro de texto y libro infantil eran 7.17 y 3.40 respectivamente. La media y desviación típica para el grupo que empleó únicamente el libro de texto eran 3.80 y 2.07 respectivamente. Las pruebas *a posteriori* (SNK) indicaron que los dos grupos instruidos con el libro infantil diferían significativamente del grupo instruido sólo con el libro de texto.

TABLA 3: Resultados del análisis de varianza de una vía para el subtest que medía el contenido del libro infantil

Fuente	gl	Suma de cuadrados	Media al cuadrado	F-valor	P-valor
Inter	2	253.3526	126.6763	10.2272	.0002
Intra	54	668.8579	12.3863		
Total	56	922.2105			

Fuente: Mc Kinney, C. W., y Jones, H. J. (1993), «Effects of a children's book and a traditional textbook fifth-grade students' achievement and attitudes toward social studies», *Journal of Research Development in Education*, 27 (1), 56-62.

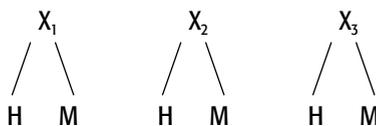
Las **comparaciones planificadas** se parecen a las pruebas *a posteriori* en que se comparan parejas de medias, pero utiliza comparaciones específicas que se identifican antes de completar la investigación. Normalmente se incluyen en el análisis unas cuantas de todas las posibles parejas de medias, aquellas que tienen

un interés especial. Como las parejas se identifican antes, y no tras el ANOVA, puede también hablarse de *pruebas a priori*.

Análisis factorial de varianza

Hemos presentado el ANOVA de un factor como un procedimiento que se utiliza con una variable independiente y dos o más niveles identificados de esta variable. Sin embargo, es frecuente manejar más de una variable independiente en un estudio. De hecho, a menudo es deseable contar con varias variables independientes porque el análisis aportará más información. Por ejemplo, si un grupo de investigadores investiga la eficacia de tres currículos de lectura, probablemente utilizarían una ANOVA 1×3 para examinar la hipótesis nula de que no hay ninguna diferencia de rendimiento entre ninguno de los tres grupos (es decir, $\bar{X}_1 = \bar{X}_2 = \bar{X}_3$). Si los investigadores estuvieran también interesados en conocer si existe un rendimiento diferencial entre niños y niñas, el género se convertiría en una segunda variable independiente. Como cada grupo de lectura de mujeres y hombres es analizado por separado, ahora hay seis grupos. Si X es el currículum de lectura y H/M es el género, entonces los seis grupos son: X_1M ; X_1F ; X_2M ; X_2F ; X_3M y X_3F . Esta situación aparece representada en el diagrama que se presenta a continuación.

Primera variable independiente:
grupos de currículum de lectura



Segunda variable independiente: género

Otro modo de ilustrar el estudio es situar cada variable independiente en un lado de un rectángulo, del siguiente modo:

		Currículum de lectura		
		X_1	X_2	X_3
Género	H			
	M			

Por consiguiente, en esta situación hipotética, hay dos variables independientes analizadas simultáneamente y una variable dependiente (rendimiento). El procedimiento estadístico que se utilizaría para analizar los resultados sería un ANOVA de dos vías o factores (*dos vías* a causa de las dos variables independientes). Como *factor* es otra palabra para la variable independiente *factorial*

significa que hay más de una variable independiente. Por lo tanto, un **ANOVA factorial** es un término genérico que significa que se analizan dos o más variables independientes juntas. El término más específico, como ANOVA de dos o tres vías, nos dice el número exacto de variables independientes. Los investigadores pueden precisar, aún más, al indicar cuál es el análisis incluyendo los niveles de cada variable independiente. Como se señalaba anteriormente el término *niveles* hace referencias a los subgrupos o categorías de cada variable independiente. En el ejemplo mencionado antes, el *currículum de lectura* tiene tres niveles y *género* tiene dos. Los niveles pueden mostrarse mediante números tras la abreviatura ANOVA. En nuestro ejemplo de lectura, se trata de un ANOVA 2×3 . (Para un ANOVA con tres factores tendría que haber tres números, como en un ANOVA $2 \times 2 \times 3$. Esto significa que hay dos niveles en dos de las variables y tres niveles en una variable). Haciendo uso de esta notación, un investigador puede comunicar mucha información de manera concisa. Normalmente, el número de factores es dos o tres y el número de niveles puede ser cualquier número que sea mayor que uno (aunque raramente supera los 10). En el anterior ejemplo hipotético, si hubiera cuatro currículos de lectura y el investigador estuviera interesado en conocer la manera en la que cada currículum afecta a los alumnos de rendimiento alto, medio y bajo, entonces el análisis resultante sería un ANOVA 3×4 . Se trata todavía de un ANOVA con dos factores, pero el número de niveles es diferente. Esta situación puede ilustrarse en el diagrama siguiente:

		Currículum de lectura			
		X_1	X_2	X_3	X_4
Nivel de capacidad	Alto				
	Medio				
	Bajo				

El extracto 9.5 es un ejemplo de la manera en la que los investigadores hacen referencia a diseños factoriales con dos variables y de cómo se presentan los resultados.

Aquí tenemos otro ejemplo hipotético para aclarar las pruebas de significación que resultan de un ANOVA con dos factores: un profesor está interesado en si las técnicas específicas para ayudar a recordar son efectivas en la mejora del rendimiento de alumnos con alto y bajo nivel de ansiedad. El profesor ha desarrollado dos técnicas, una con ayudas mnemotécnicas y la otra con práctica distribuida para la memorización del material. El profesor tiene también un grupo de control. Por lo tanto, hay dos variables independientes, una con dos niveles (ansiedad: alta y baja), la otra con tres niveles (técnicas de tratamiento: ayuda

mnemotécnica, práctica distribuida y grupo de control). Esto constituiría un diseño ANOVA 2×3 y aparece representado en la figura 9.4.

Dentro de cada cuadrado del diagrama está la media de ese grupo. Estos cuadrados se denominan *celdas*. Por consiguiente, el número 50 que se encuentra en la celda de arriba a la derecha (\bar{X}_5) se refiere a la media de los sujetos con un alto nivel de ansiedad que servían como grupo de control. El ANOVA 2×3 examina tres hipótesis nulas: que no hay ninguna diferencia entre los alumnos con alta y baja ansiedad; que no hay ninguna diferencia entre las condiciones de tratamiento y de control, y que no hay ninguna interacción entre los dos factores (la *interacción* aparece definida en el próximo párrafo). Las dos primeras hipótesis son parecidas en la interpretación de ANOVAs de una vía. Éstas indican al investigador si se dan diferencias en cada uno de los factores independientemente de los otros. Esto se denomina *efectos principales* o *simples* en una ANOVA factorial.

EXTRACTO 9.5

ANOVA FACTORIAL

Utilizamos un análisis de varianza 2×2 (programa: instrucción directa frente a Instrucción tradicional \times movilidad: estudiantes que se desplazan frente a estudiantes estables) para analizar cada puntuación total y las puntuaciones de cada componente.

Para los efectos principales, los estudiantes con instrucción directa obtuvieron resultados significativamente más altos que los estudiantes tradicionales en cálculos matemáticos. Asimismo, los estudiantes estables obtuvieron resultados significativamente más altos que los estudiantes que se desplazan en lectura, uso de la palabra, vocabulario, comprensión, y en la puntuación total de la batería. No hubo diferencias significativas en los resultados totales del lenguaje ni en dos de sus componentes, mecánica y expresión, así como tampoco la hubo en los resultados matemáticos totales.

Hubo también tres interacciones significativas relacionadas con el vocabulario, deletreo, y los conceptos matemáticos. Una prueba *a posteriori* utilizando el HSD reveló que los estudiantes estables tradicionales tenían resultados significativamente más altos que los estudiantes tradicionales móviles en deletreo y vocabulario. En los conceptos matemáticos y su aplicación, los estudiantes tradicionales estables tenían resultados significativamente más altos que los estudiantes tradicionales que se desplazan y los estudiantes estables con instrucciones directas.

Un examen de las medias de todos los grupos revela un patrón parecido de diferencias significativas. El patrón muestra un nivel parecido de rendimiento tanto para los estudiantes estables como para los que se desplazan con instrucciones directas. Las medias estables tradicionales son similares en los dos grupos de instrucción directa. Sin embargo, las medias de los estudiantes que se desplazan tradicionales son más bajas que las de sus compañeros estables y los dos grupos de instrucción directa.

Fuente: Brent, G. y DiObilda, N. (1993), «Effects of curriculum alignment versus direct instruction on urban children», *Journal of Educational Research*, 86, 335.

FIGURA 9.4: ANOVA HIPOTÉTICO 2 × 3

		Técnicas de tratamiento			
		Ayudas mnemotécnicas	Práctica distribuida	Control	
Ansiedad	Alta	$\bar{X}_1 = 62$	$\bar{X}_3 = 50$	$\bar{X}_5 = 50$	$\bar{X}_7 = 54$
	Baja	$\bar{X}_2 = 44$	$\bar{X}_4 = 59$	$\bar{X}_6 = 59$	$\bar{X}_8 = 54$
		$\bar{X}_9 \cong 53$	$\bar{X}_{10} \cong 55$	$\bar{X}_{11} \cong 55$	

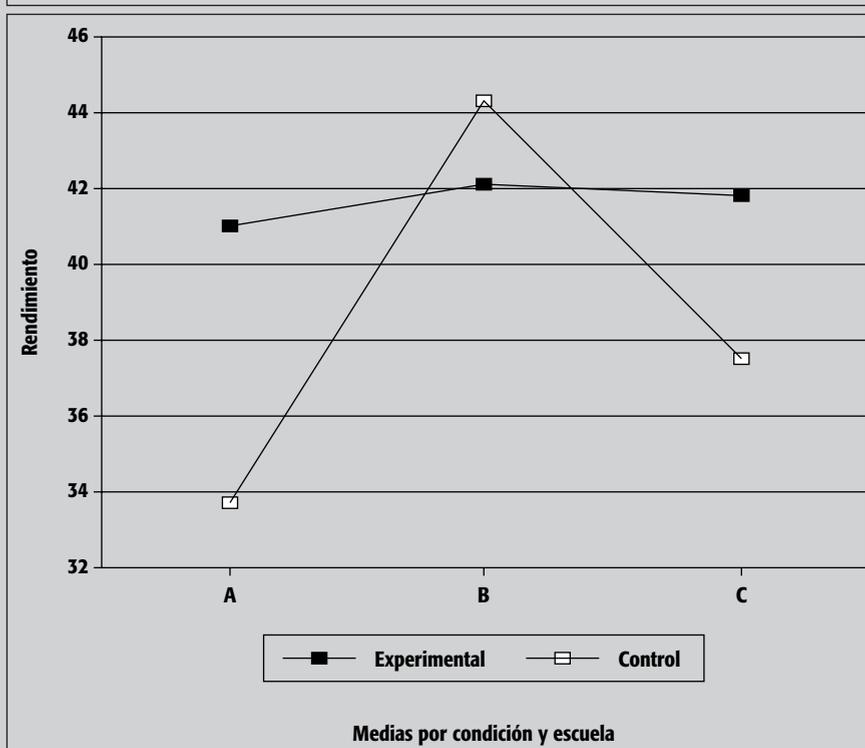
Hay un efecto principal (no necesariamente significativo) para la ansiedad y otro efecto principal para las técnicas de tratamiento. Al calcular el ANOVA 2 × 3 habrá un valor F diferente para cada efecto principal, con niveles correspondientes de significación. En nuestro ejemplo, el efecto principal para la ansiedad se somete a examen comparando \bar{X}_7 con \bar{X}_8 . Esta fila de medias no tiene en cuenta la influencia de las técnicas y sólo consideran la ansiedad. Como \bar{X}_7 y \bar{X}_8 son 54, la hipótesis nula de que $\bar{X}_7 = \bar{X}_8$ no se rechazaría; es decir, examinando sólo la ansiedad no hay ninguna diferencia de rendimiento entre los estudiantes con una ansiedad alta y baja. Para el efecto principal de la técnica, se comparan \bar{X}_9 , \bar{X}_{10} y \bar{X}_{11} . Una vez más, parece que hay muy poca diferencia en el rendimiento entre los tres grupos de técnicas. Las dos primeras hipótesis nulas podrían haber sido sometidas a examen usando ANOVAs de un factor independientes, pero el ANOVA 2 × 3 es más preciso, más potente en la detección de diferencias y más parsimonioso que el empleo de dos ANOVAs de un factor. Además, el ANOVA 2 × 3 permite al investigador comprobar la tercera hipótesis nula. Esta hipótesis está relacionada con lo que se conoce como una *interacción* entre las variables independientes.

Una **interacción** es el efecto de las variables independientes juntas; es decir, que el impacto de un factor en la medida dependiente varía con el nivel de un segundo factor. Dicho con otras palabras, es el efecto conjunto de las variables independientes sobre la variable dependiente. Una interacción es evidente si las diferencias entre los niveles de una variable independiente son inconsistentes de un nivel a otro de la otra variable independiente. En otras palabras, existe una interacción si el efecto de una variable difiere a lo largo de los diferentes niveles de la segunda variable. En nuestro ejemplo, si observamos la diferencia entre \bar{X}_1 y \bar{X}_2 ($62 - 44 = 18$) y comparamos esta diferencia con $\bar{X}_3 - \bar{X}_4$ (-9) y $\bar{X}_5 - \bar{X}_6$ (-9), encontramos que existe una gran diferencia entre nivel de ansiedad alto o bajo si nos situamos en cada una de las técnicas de tratamiento. Esta es la evidencia visual de una interacción. Estadísticamente, se presenta un valor F para la interacción con su nivel correspondiente de significación. Esta prueba estadística se denomina el efecto de interacción.

EXTRACTO 9.6 INTERACCIÓN ANOVA

El análisis de varianza mostró un efecto altamente significativo de la condición experimental ($F = 10.59, p < .01$), un efecto significativo de la escuela ($F = 3.35, p = <.05$) y una interacción entre la condición y la escuela ($F = 5.01, p < .05$)... El efecto motivacional era el mismo para chicos y chicas y constante a lo largo de los cursos, pero difería entre escuelas. La figura 1 refleja un efecto muy grande en la escuela A; un gran efecto en la escuela C, y el grupo de control, de alguna manera, mayor que el grupo experimental en la escuela B.

FIGURA 1: MEDIAS POR CONDICIÓN Y ESCUELA



Fuente: Brown, S. M. y Walberg, H. J. (1993), «Motivational effects of test scores of elementary students», *Journal of Educational Research*, 86, 134, 135.

Ahora está claro cómo un ANOVA factorial puede facilitar más información que distintos ANOVAs con un único factor. En nuestro ejemplo, es evidente que las técnicas de tratamiento suponen un cambio para los alumnos con un nivel de

ansiedad alto o bajo. Los alumnos con un alto nivel de ansiedad obtienen mejores resultados con las ayudas mnemotécnicas que con la práctica distribuida o sin tratamiento alguno, mientras que los alumnos con un bajo nivel de ansiedad obtienen mejores resultados con la práctica distribuida o sin tratamiento que con las ayudas mnemotécnicas. Este resultado sería estadísticamente significativo a pesar de que ninguno de los efectos principales era significativo por sí mismo. En los extractos 9.5 y 9.6, los autores presentan interacciones significativas.

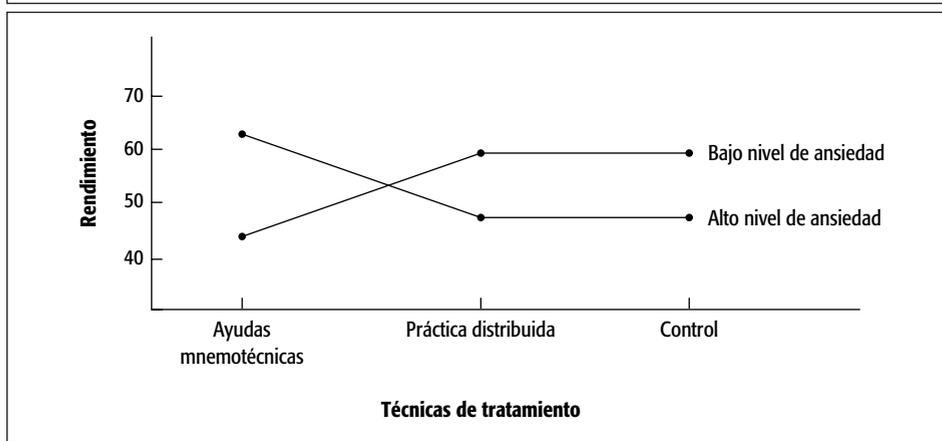
Es frecuente presentar un gráfico de la interacción cuando es significativa. El gráfico muestra la naturaleza de una interacción de manera más clara de lo que lo hacen las muestras de las celdas. El gráfico se construye colocando valores para la variable dependiente a lo largo del eje vertical (ordenada) y los niveles de una variable independiente sobre el eje horizontal (abscisa); todas las medias de las celdas están representadas en el gráfico e identificadas con la segunda variable independiente. Para nuestro ejemplo hipotético, la interacción aparece representada en la figura 9.5. En ésta, se utilizan líneas para conectar las medias de las celdas. Si las líneas son paralelas, entonces no hay interacción. Si las líneas son horizontales, se dice que la interacción no es ordinal.

Análisis de covarianza

El **análisis de covarianza** (ANCOVA) es un procedimiento estadístico que se utiliza en casos parecidos a los que se utiliza un ANOVA de una vía o uno factorial. Un ANCOVA tiene dos objetivos principales: ajustar estadísticamente las diferencias iniciales de grupo en una o más variables relacionadas con la variable dependiente pero no controladas y aumentar la probabilidad de encontrar una diferencia significativa entre las medias de los grupos.

Para el primer objetivo, consideremos el siguiente ejemplo. Un investigador utiliza dos clases para investigar si es más efectiva la enseñanza en equipo o la individualizada. Basándose en un pretest, el investigador sabe que una clase tiene un mayor conocimiento de la variable dependiente (rendimiento en matemáticas) que el otro grupo (por ejemplo, la media del pretest del grupo cooperativo es 12 y la del grupo individualizado es 10). Si realiza un postest y encuentra que la media del grupo cooperativo es 24 y la del grupo individualizado 20, el investigador podría inclinarse por concluir que el grupo cooperativo ha rendido más que el individualizado. Es probable que esto ocurra si se ignoran los resultados del pretest. Un enfoque alternativo sería observar las puntuaciones de ganancia entre el pretest y el postest y utilizar una prueba t para determinar si estas puntuaciones son significativamente diferentes. Este enfoque resultaría al comparar 12 (24 - 12) con 10 (20 - 10). Aunque este enfoque es teóricamente mejor que no utilizar los resultados del pretest, por razones que superan el contenido de este libro existen problemas técnicos al comparar las puntuaciones de ganancia. El

FIGURA 9.5: INTERACCIÓN PARA ESTUDIO HIPOTÉTICO



mejor método para analizar los datos en esta circunstancia es el empleo de un ANCOVA. El ANCOVA ajustaría estadísticamente las puntuaciones del postest mediante las diferencias que existían entre los grupos en el pretest. En este ejemplo, la puntuación postest del grupo cooperativo disminuiría un punto, porque la media de este grupo era más alta por un punto que la media de ambos grupos en el pretest. Igualmente, como la media del grupo individualizado en el pretest es un punto más bajo que la media de los dos pretests, el resultado del postest de 20 aumentaría en un punto a 21. De este modo, en vez de comparar 20 con 24, el ANCOVA compararía 21 con 23.

La variable que se utiliza en el ANCOVA para ajustar las puntuaciones (en el ejemplo anterior el pretest) se denomina *variable concomitante o covariante*. Las covariantes son, a menudo, puntuaciones de pretests o resultados de pruebas de rendimiento, actitud o aptitud que estarían relacionados con la variable dependiente. Las puntuaciones de CI y las de tests estandarizados de rendimiento, por ejemplo, son utilizadas con frecuencia como covariantes.

El segundo objetivo del análisis de covarianza es aumentar lo que se conoce como la «potencia» de la prueba estadística para encontrar diferencias entre grupos. La explicación detallada del concepto de potencia va más allá del contenido de este libro. Resumidamente, la potencia es la probabilidad de detectar una diferencia significativa. Es útil aumentar la potencia cuando el tamaño muestral es pequeño o cuando el investigador tiene razones para creer que las diferencias entre los grupos serán pequeñas.

Se puede emplear un ANCOVA en muchas situaciones: con dos grupos y una variable independiente en lugar de una prueba *t*, con una variable independiente que tiene más de dos grupos en lugar del ANOVA de una vía y con el análisis de varianza factorial. Se pueden usar también en investigación más de una cova-

riante en un único procedimiento ANCOVA. La presentación de un ANCOVA es muy parecida a la presentación del ANOVA. El extracto 9.7 muestra cómo se utilizó el análisis de covarianza en un estudio real para ajustar diferencias iniciales de grupo.

Dado que el ANCOVA se utiliza a menudo con grupos intactos, sin asignación aleatoria, hay que advertir que la interpretación de los resultados debería sopesar la posibilidad de que otras variables no controladas ni medidas estén también relacionadas con la variable dependiente y que, por lo tanto, puedan tener un efecto sobre ésta. En otras palabras, aunque puede obtenerse el ajuste estadístico del efecto de la covariable, el investigador no puede llegar a la conclusión de que los grupos son iguales en el sentido de la asignación aleatoria.

EXTRACTO 9.7

USO DE ANCOVA PARA EL AJUSTE DE LAS DIFERENCIAS DE GRUPO

Se realizaron 14 ANCOVAs que tenían grupo (experimental orientado a la actividad, experimental orientado a la discusión y control) como variable independiente... Para los 14 ANCOVAs se usaron como covariantes las puntuaciones grupales del pretest de asistencia al colegio, calificación media, número de expedientes disciplinarios, puntuación total de la escala Piers-Harris y el «cluster» Satisfacción y Felicidad de la misma escala. El uso de ANCOVAs permitió a los investigadores el ajuste de todas las diferencias pretratamiento que existían entre los grupos experimentales y el grupo de control.

Fuente: Page, R. C. y Chandler, J. (1994), «Effects of group counseling on ninth-grade at-risk students», *Journal of Mental Health Counseling*, 16, 346.

PRUEBAS NO PARAMÉTRICAS

En la presentación de pruebas estadísticas realizada hasta ahora hemos tratado procedimientos que utilizan estadísticas muestrales para estimar características de una población. Estas características de la población se llaman parámetros y se hace referencia a los procedimientos estadísticos como procedimientos **paramétricos**. Además, las estadísticas paramétricas se utilizan cuando el investigador puede asumir que la población se distribuye normalmente, que existe homogeneidad de varianzas entre los diferentes grupos y posee los datos de intervalo o razón de una escala.

Siempre que se den las condiciones en las que se basan las estadísticas paramétricas, el investigador utilizará una prueba *t*, un ANOVA, un ANCOVA o algún otro procedimiento paramétrico. Si no se cumplen estas condiciones —es decir, si los datos no son de intervalo o razón o no tienen una distribución

normal— el investigador debería considerar utilizar una prueba **no paramétrica** análoga a la prueba paramétrica. Para la mayoría de los procedimientos paramétricos existe una prueba no paramétrica correspondiente que puede utilizarse. La interpretación de los resultados es parecida en los dos tipos de pruebas. Se diferencian en la ecuación calculada y en las tablas para determinar el nivel de significación de los resultados. Los dos procedimientos comprueban una hipótesis y presentan un nivel de significación para rechazar la hipótesis nula. Sin embargo, frente a las pruebas paramétricas, las pruebas no paramétricas no ponen a prueba hipótesis sobre las características de una población. En su lugar, los procedimientos no paramétricos examinan hipótesis sobre las relaciones entre variables categóricas, formas de las distribuciones y normalidad de una distribución. Mientras los procedimientos paramétricos utilizan medias, las técnicas no paramétricas están relacionadas con frecuencias, porcentajes y proporciones. En general, las pruebas paramétricas son más potentes en la detección de diferencias significativas y se utilizan con frecuencia, incluso, cuando no se cumple ninguna condición.

La tabla 9.2 proporciona los nombres de las pruebas no paramétricas que son análogas a los pruebas paramétricas que acabamos de comentar.

TABLA 9.2: PROCEDIMIENTOS PARAMÉTRICOS Y NO PARAMÉTRICOS	
Paramétrico	No paramétrico
Prueba t para muestras independientes	Prueba de la mediana Prueba U de Mann-Whitney
Prueba t para muestras dependientes	Prueba de los signos Prueba de Kruskal-Wallis
ANOVA de una vía	Prueba de la mediana Prueba de los rangos signados para datos pareados o prueba de Wilcoxon

Para más información sobre pruebas no paramétricas, ver Siegel (1956) Marascuilo y Mc Sweeney (1977) y Gibbons (1993).

La prueba **chi-cuadrado** es un procedimiento no paramétrico común que se utiliza cuando los datos están en forma nominal. Esta prueba es una manera de responder a las preguntas sobre asociación o relación basada en las frecuencias de observaciones en distintas categorías. Las frecuencias pueden estar en casi cualquier forma —personas, objetos, votos— y, simplemente, se cuentan dentro de cada categoría. De este modo, el investigador forma las categorías y, a continuación, cuenta la frecuencia de observaciones o sucesos dentro de cada categoría.

	Novatos	Alumnos de segundo año	Alumnos de tercer año	Alumnos de último año
Observado	30	25	15	30
Esperado	25	25	25	25

En la prueba de chi-cuadrado para una muestra, el investigador maneja una variable independiente que está dividida en dos o más categorías. Por ejemplo, un director de instituto puede estar interesado en el número de nuevos alumnos, alumnos de segundo año, alumnos de tercer año y alumnos de último curso que asistían al centro de orientación o, en otras palabras, la relación entre año de instituto y utilización de los servicios de orientación. La variable independiente es año de instituto, con cuatro categorías. El investigador podría seleccionar una muestra aleatoria de cincuenta alumnos de cada categoría y registrar en cada categoría el número de los que asistían al centro de orientación. La prueba estadística compara las frecuencias recogidas u observadas con algunas frecuencias teóricas o esperadas. En nuestro ejemplo, el director del instituto podría esperar que las frecuencias de cada categoría fueran las mismas. Entonces, la hipótesis nula puesta a prueba sería que no hay ninguna diferencia en el número de alumnos que asistían al centro de orientación entre las cuatro categorías.

La siguiente tabla ilustra este ejemplo.

Para obtener el nivel de significación el investigador calcula una fórmula para obtener un valor chi-cuadrado (χ^2), utiliza los grados de libertad adecuados y consulta una tabla chi-cuadrado (ver Apéndice B) para determinar el nivel de significación al rechazar la hipótesis nula. En nuestro ejemplo, la prueba sería significativa, mostrando que los alumnos de primer y último curso asisten más al centro de orientación más que los de segundo y tercer año y que éstos últimos son los que menos asisten de todas las clases.

Si el investigador tiene más de un variable independiente, para analizar los datos puede utilizar la **prueba chi-cuadrado para muestras independientes del chi-cuadrado o tabla de contingencia**. En el ejemplo anterior, si el director estuviera también interesado en conocer las diferencias entre hombres y mujeres en cada nivel de clase, entonces el análisis sería como una ANOVA factorial. En este caso sería una tabla de contingencia de 2×4 .

Se puede dar diversos usos a la prueba chi-cuadrado. Puede utilizarse en investigación de actitudes si el investigador clasifica las respuestas como favorables o no favorables; con alumnos de capacidad alta, media y baja que exhiben su capacidad durante una tarea; en especial investigación educativa con frecuencias de conducta adecuada y muchos otros problemas. Los investigadores pueden presentar una medida genérica de relación con los resultados chi-cua-

drado. Estas medidas se denominarían **coeficiente phi o coeficiente de contingencia** y serían interpretadas aproximadamente del mismo modo que un coeficiente de correlación producto-momento de Pearson. El extracto 9.8 es un ejemplo de presentación de resultados de una prueba chi-cuadrado en un estudio real.

ANÁLISIS MULTIVARIADO

Nuestra presentación de la Estadística inferencial sería incompleta si no introdujéramos el análisis *multivariado*. Los científicos sociales saben, desde hace muchos años, que el comportamiento humano puede entenderse mejor en situaciones complejas examinando muchas variables al mismo tiempo y no ocupándose de una o dos variables en cada estudio. Los procedimientos estadísticos para analizar muchas variables coincidentes en el tiempo han estado disponibles durante muchos años, pero los investigadores sólo han podido utilizar estos procedimientos desde la utilización de los ordenadores. Actualmente, estos procedimientos estadísticos más complejos se presentan con frecuencia en las revistas científicas.

EXTRACTO 9.8 **CHI-CUADRADO**

Resultados

En la pregunta acerca de su opinión general sobre la repetición de curso, se les pidió a los padres, profesores y directores que señalaran uno de los siguientes términos: «nunca», «raramente», «ocasionalmente», «normalmente» o «siempre» como respuesta a la pregunta, «los niños deberían repetir curso si no alcanzan los requisitos necesarios del curso». Cuando se examinaron las respuestas de padres, profesores y directores mediante un análisis de chi-cuadrado, los contrastes alcanzaron el nivel de significación .0001. Como se muestra en la tabla 1, la diferencia más obvia residía en la tendencia de los padres de señalar las respuestas extremas de «nunca» y «siempre» con más frecuencia que los profesores o los directores.

A los tres grupos se les solicitó también que seleccionaran, entre las razones siguientes, las que consideraban válidas como criterio para la repetición de curso: (a) la falta crónica de asistencia, (b) la petición de los padres, (c) la inmadurez emocional, (d) el fracaso académico por otras razones que la falta de destrezas básicas y (e) falta de destrezas básicas. Como se muestra en la tabla 1, las opiniones de los padres, profesores y directores era significativamente diferente acerca de las ausencias excesivas ($p < .0001$); madurez emocional ($p < .0001$); fracaso académico por otras razones que la falta de habilidades básicas ($p < .0006$) y la falta de habilidades básicas ($p < .0001$). La última pregunta igual en los tres cuestionarios preguntaba a los padres, profesores y directores

quién debería tener la última palabra en la repetición de curso del niño. Era evidente una diferencia significativa en sus opiniones ($p < .0001$) (ver tabla 1).

TABLA 1: Comparaciones entre padres, profesores y directores

	Opiniones representativas de sus puntos de vista sobre la repetición de curso					N
	Nunca	Excepcionalmente	Ocasionalmente	Con frecuencia	Siempre	
Padres	3,6%	13.0	23.9	37.0	22.6	1,063
Profesores	.7	9.7	24.8	52.4	12.4	145
Directores	.0	8.6	17.1	71.4	2.9	35

($\chi^2 = 35.474$; $p < .0001$)

	Razones apropiadas para repetir					
	Ausencias excesivas		Peticiónes de los padres		Inmadurez emocional	
	Sí	No	Sí	No	Sí	No
Padres	13.6%	86.3	15.2	84.8	18.6	81.4
Profesores	39.3	60.7	15.9	84.1	53.1	46.9
Directores	68.6	31.4	20.0	80.0	54.3	45.7

($\chi^2 = 119.109$; $p < .0001$) ($\chi^2 = 103.933$; $p < .0001$)

	Razones apropiadas para repetir					
	Fracaso académico		Falta de habilidades básicas		Otras razones	
	Sí	No	Sí	No	Sí	No
Padres	29.6%	70.3	62.6	37.4	5.9	94.1
Profesores	36.6	63.4	85.4	13.9	—	—
Directores	57.1	42.9	94.3	5.7	—	—

($\chi^2 = 14.271$; $p < .0006$) ($\chi^2 = 51.451$; $p < .0001$)

	Palabra final sobre la decisión de repetir								
	(P)	(D)	(Pa)	(N)					
	Profesor	Director	Padre	Niño	P-D	P-D-Pa	P-Pa	D-Pa	Resto
Padres	47.8%	14.6	19.6	.4	3.0	3.2	9.2	1.5	.7
Profesores	66.2	11.7	3.4	.0	13.1	3.4	.7	.0	1.4
Directores	22.9	54.3	2.9	.0	11.4	5.7	.0	.0	2.9

($\chi^2 = 128.481$; $p < .0001$)

La muestra estaba compuesta por 1.063 padres, 145 profesores y 35 directores.

Fuente: Byrnes, D. y Yamamoto, K. (1986), «Views on grade repetition», *Journal of Research and Development in Education*.

El término **multivariado** hace referencia a los métodos que investigan patrones entre muchas variables o a estudios que implican dos o más variables dependientes relacionadas para cada sujeto. Muchos investigadores se referirán a la regresión múltiple como un procedimiento multivariado, mientras que otros investigadores utilizan el término multivariado para describir los análisis en cualquier estudio de muchas variables. En cambio, los diseños que emplean pruebas *t*, ANOVAs y ANCOVAs con una única variable dependiente serían claramente clasificados como *univariados*.

Todos los procedimientos estadísticos comentados hasta este momento contaban sólo con una variable dependiente. Sin embargo, hay muchos casos en los que el investigador está interesado en más de una variable dependiente. Por ejemplo, si un investigador está estudiando actitudes hacia la ciencia, serían interesantes muchos aspectos de una actitud general hacia la ciencia, tales como disfrutar de la ciencia y valorarla, estudiarla tanto en química como en biología, tanto para la disección como para excursiones científicas, etc. De hecho, muchas pruebas de actitud tienen subescalas que reflejan de manera más específica y precisa los sentimientos y las impresiones de lo que puede hacerle una puntuación general. El investigador podría combinar todos estos diferentes aspectos y considerar la actitud como una disposición general, pero es mejor examinar cada aspecto por separado. ¿Por qué no utilizar un análisis univariado por separado para cada variable dependiente? Es decir, ¿por qué no calcular tantas ANOVAs o pruebas *t* como variables dependientes? La razón es que si las variables dependientes están correlacionadas, el uso de análisis univariados independientes aumentará la probabilidad de encontrar una diferencia simplemente porque se han empleado tantas pruebas. Se asemeja a la razón por la que en el ANOVA usamos pruebas *a posteriori* en vez de muchas pruebas *t*. Los análisis multivariados son también más concretos, es decir más directos, más rápidos y con menos cálculos separados.

TABLA 9.3: ANÁLOGOS MULTIVARIADOS

Pruebas univariadas	Pruebas multivariadas
Prueba <i>t</i>	T ² de Hotelling
ANOVA	MANOVA (análisis multivariado de varianza)
ANCOVA	MANCOVA

Aunque el cálculo y la interpretación de exámenes multivariados resultan bastante complejos, el principio básico de rechazo de la hipótesis nula en algunos niveles de significación es el mismo que para toda la Estadística inferencial. La

diferencia es que se consideran todas las variables dependientes juntas en un único análisis. Para la mayoría de los procedimientos comentados anteriormente que tienen una variable dependiente, se puede emplear un análogo multivariado cuando hay la misma variable o variables independiente(s) pero más de una variable dependiente. La tabla 9.3 resume algunas pruebas multivariadas que se corresponden con los procedimientos utilizados con una variable dependiente.

El extracto 9.9 ilustra la manera en la que un investigador utilizó un procedimiento multivariado, un MANOVA, para analizar los datos.

EXTRACTO 9.9 **ANOVA MULTIVARIADO**

Las interconexiones entre las puntuaciones de las cuatro escalas de cuestionarios previos y posteriores aparecen ilustradas en la tabla III. Estas interconexiones muestran que hay una relación positiva, pero modesta, entre las puntuaciones de las dos subescalas RSA (positiva frente a negativa). La medida de afecto hacia la docencia está relacionado de manera positiva con las puntuaciones de las dos subescalas RSA, pero más con el resultado positivo que con el resultado negativo. Sin embargo, la medida de la autoestima docente está negativamente relacionada con las puntuaciones de las dos subescalas RSA y la de afecto hacia la docencia. Además, las relaciones son relativamente fuertes. Estas interconexiones significativas entre este conjunto de medidas dependientes sugiere la necesidad de procedimientos de análisis multivariados. De este modo, para comparar las diferencias entre los grupos de profesores, se realizaron análisis multivariados de la varianza (MANOVAs), primero sobre las medidas del cuestionario preliminar y, a continuación, sobre las del cuestionario posterior. En estos análisis se consideró el grupo de los profesores como una variable independiente que tenía cuatro niveles y las puntuaciones de las escalas incluidas en los cuestionarios se consideraron cuatro medidas dependientes interrelacionadas. Los procedimientos MANOVA no suponen restricción alguna en la cantidad de sujetos en los subgrupos, de ahí que sean apropiados incluso cuando, como en este caso, las cantidades son desiguales y desproporcionadas (Finn y Wattsson, 1978, p.32).

Los resultados a partir de los análisis multivariados de los cuestionarios preliminares y posteriores aparecen representados en la tabla IV. Estos análisis mostraron que inicialmente no había diferencias estadísticamente significativas entre los grupos de los profesores. Todos los grupos eran comparables en cuanto a valoraciones de responsabilidad para resultados positivos, la responsabilidad para resultados negativos y el afecto hacia la docencia. Aquellos profesores que no utilizaron las nuevas estrategias parecían expresar de algún modo una mayor confianza en sus capacidades para enseñar que los otros grupos de profesores, pero esta diferencia no era estadísticamente significativa.

Los análisis de las medidas de los cuestionarios posteriores indicaron que las diferencias resultantes entre los grupos de los profesores eran estadísticamente significativas. Se realizaron comparaciones *a posteriori* calculando intervalos de confianza Scheffé de .95. Estas comparaciones revelaron que los profesores que habían experimentado un

cambio positivo en los resultados de aprendizaje de sus alumnos se sentían más responsables tanto de los resultados negativos como de los positivos de sus alumnos, y expresaron actitudes más positivas hacia la enseñanza que los otros grupos de profesores. Sin embargo, al contrario de lo que habíamos considerado como hipótesis, estos profesores expresaron una menor confianza en sus capacidades para enseñar.

Curiosamente, los análisis *a posteriori* mostraron también diferencias estadísticamente significativas para aquellos profesores que no utilizaron las nuevas estrategias. Los profesores de este grupo expresaron actitudes más negativas hacia la enseñanza que los otros grupos de profesores, pero también expresaron mucha más confianza en sus técnicas docentes.

TABLA III. Interconexiones entre medidas de variables afectivas seleccionadas

Variable	RSA positivo	RSA negativo	Afecto hacia la docencia	Autoestima como docente
RSA positivo		0.233	0.485*	-0.510*
RSA negativo	0.341*		0.330*	-0.062
Afecto a la enseñanza	0.427*	0.251*		-0.638*
Autoestima como docente	-0.481*	-0.173	-0.602*	

Nota: N = 17

* $p < 0.01$

TABLA IV. Resumen de análisis multivariado de la varianza

Pruebas de significación		F univariadas				
Fuentes de variación	gl	F multivariada	R+	R-	Afecto hacia la docencia	Autoestima docente
Constante	1	—	—	—	—	—
Grupo (cuestionario preliminar)	3	0.97	1.08	0.08	0.52	2.52
Grupo (cuestionario posterior)	(3)	15.07*	36.34*	10.92*	18.99*	29.17*
Media cuadrática simple						
Total	117	Cuestionario previo	95.88	97.57	34.64	116.01
		Cuestionario posterior	77.24	94.37	37.27	82.21

* $p < 0.001$

Nota: Guskey, T. R. (1984), «The influence of change in instructional effectiveness characteristics of teachers», *American Educational Research Journal*, 21, 250-252.

RESUMEN

Este capítulo ha presentado la lógica de la Estadística inferencial y ha descrito algunos de los procedimientos estadísticos más frecuentes empleados por los investigadores para analizar los datos. Los siguientes apartados resumen los conceptos presentados.

1. La Estadística inferencial se utiliza para hacer deducciones basadas en aspectos medidos de una muestra sobre las características de una población.
2. Durante la realización de la investigación, la probabilidad es una preocupación a causa del error implicado en el muestreo y la medida.
3. Las estadísticas muestrales representan estimaciones imperfectas de la población.
4. La Estadística inferencial estima la probabilidad de que las características de la población se encuentren en un cierto rango de valores. Los niveles de confianza se presentan con frecuencia.
5. La hipótesis nula se utiliza para poner a prueba la presuposición de que no hay diferencia alguna entre los valores de la población.
6. Los investigadores intentan rechazar la hipótesis nula utilizando pruebas estadísticas inferenciales para indicar la probabilidad de equivocarse al rechazar la hipótesis nula.
7. El nivel de significación de una prueba estadística informa al investigador de la probabilidad de cometer un error tipo I al rechazar la hipótesis nula.
8. La mayoría de los investigadores usan un nivel p de .05 o menos ($< .05$) para indicar la significación estadística.
9. Las medidas de magnitud del efecto, tales como el tamaño de efecto, indican la magnitud de las diferencias y las relaciones.
10. La prueba t se utiliza para comparar dos medias y , según la naturaleza de la investigación, usa una ecuación para muestras independientes o una ecuación para muestras dependientes para calcular el valor t .
11. El análisis de varianza se utiliza para comparar dos o más medias y presenta el estadístico ϕ con los niveles p correspondientes.
12. El ANOVA de una vía o factor examina la diferencia entre los niveles de una variable independiente, mientras el ANOVA factorial examina más de una variable independiente.
13. Las pruebas de comparación *a posteriori* y planificadas están diseñadas para localizar diferencias significativas entre parejas de medias en un ANOVA.
14. Una interacción se examina en el ANOVA factorial como el efecto específico de las variables independientes que actúan juntas para afectar a los resultados de la variable dependiente.

15. Las pruebas paramétricas asumen que los datos son de razón o intervalo, la homogeneidad de varianzas y una distribución normal. Las pruebas no paramétricas pueden utilizarse en casos en los que se incumplen estas condiciones y la prueba no es lo suficientemente sólida para este incumplimiento.
16. El análisis de covarianza se utiliza para ajustar las diferencias entre grupos de una variable covariante que está relacionada con la variable dependiente.
17. Chi-cuadrado es una prueba no paramétrica que aparece con frecuencia y se utiliza cuando el investigador está examinando la frecuencia de los sucesos.
18. Los análisis multivariados se utilizan en casos en los que coinciden varias variables dependientes relacionadas.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. En la investigación se hacen enunciados de probabilidad debido a...
 - a. que el número total de casos esperados es pocas veces conocido.
 - b. que existe error en el muestreo.
 - c. la fiabilidad y la validez de la medición.
 - d. que tanto b como c son correctos.
2. La hipótesis nula es un enunciado que es...
 - a. el mismo que el resultado esperado.
 - b. necesario porque sólo podemos descartarla, no probarla.
 - c. expresado de forma negativa para corresponderse con las pruebas estadísticas.
 - d. necesario en la presentación de la investigación.
3. El tamaño del efecto es útil en la determinación de...
 - a. los intervalos de confianza.
 - b. el error típico.
 - c. la magnitud.
 - d. alfa.
4. Compruebe si todas las consideraciones siguientes son ciertas sobre el nivel de significación:
 - a. $05 < .01$.
 - b. 05 es mejor que .01.
 - c. 001 es mejor que .01.
 - d. $.05 < .10$

- e. .20 es marginalmente significativo.
 - f. .01 es estadísticamente significativo.
 - g. .03 significa equivocarse al rechazar la hipótesis nula 3 veces de cada 100.
5. La prueba t se utiliza para comparar...
 - a. dos o más medias.
 - b. una media con un valor de interés establecido.
 - c. dos medias.
 - d. que tanto b como c son correctos.
 6. Otro nombre para la prueba t para muestras dependientes es...
 - a. pares relacionados.
 - b. pares emparejados.
 - c. no dependiente.
 - d. independiente.
 7. En un análisis de varianza de una vía el investigador presentaría...
 - a. un estadístico ϕ .
 - b. tres estadísticos ϕ .
 - c. Los resultados para todas las variables independientes.
 - d. Los resultados para todas las variables dependientes.
 8. El ANOVA factorial se utiliza cuando...
 - a. hay dos o más variables independientes.
 - b. hay dos o más variables dependientes.
 - c. las pruebas t no son factibles.
 - d. tanto a como c son correctas.
 9. Los procedimientos de comparaciones múltiples se utilizan porque...
 - a. son más apropiados.
 - b. reducen la probabilidad de cometer un error en las conclusiones.
 - c. es el único modo de estudiar todas las diferencias de grupo.
 - d. otros procedimientos son demasiado conservadores.
 10. Una interacción tiene lugar cuando...
 - a. los resultados de cada variable son diferentes.
 - b. el efecto de cada variable es diferente.
 - c. dos o más variables se combinan dando como resultado efectos específicos.
 - d. dos o más personas hablan entre ellas.
 11. Un ANOVA 2×4 significa...
 - a. dos variables independientes y una variable dependiente.
 - b. una variable independiente con dos niveles y una variable dependiente con cuatro niveles.
 - c. ocho variables independientes.
 - d. una variable independiente con dos niveles y otra variable independiente con cuatro niveles.

12. Como ejemplos de procedimientos estadísticos no paramétricos se incluyen:
 - a. chi-cuadrado, ANOVA, MANOVA.
 - b. MANOVA, ANCOVA.
 - c. chi-cuadrado, prueba *t*.
 - d. chi-cuadrado, test de la mediana.
13. Los procedimientos multivariados se utilizan cuando...
 - a. hay varias variables.
 - b. hay más de una variable dependiente relacionada.
 - c. las variables independientes tienen más de dos niveles.
 - d. hay múltiples pruebas por hacer.

PROBLEMAS APLICADOS

En cada uno de los ejemplos que se presenta a continuación, seleccione el procedimiento estadístico que analizaría mejor los datos.

1. Este investigador estaba interesado en cómo funcionan tres enfoques disciplinares diferentes con alumnos de quinto curso. Se asignaron profesores en prácticas de forma aleatoria a cada uno de los tres grupos. Se instruyó a cada grupo en un enfoque diferente para tratar los problemas de disciplina. A continuación, se observó a cada profesor en prácticas durante un periodo aproximado de tres semanas y se registraron los problemas de disciplina y su duración.
2. Un profesor está interesado en saber si las actitudes de sus alumnos de sexto curso frente al sexo cambiarán tras seguir un curso de educación sexual de cuatro semanas. Para evaluar el impacto del programa, el profesor mide las actitudes de los alumnos antes y después del curso.
3. El profesor del problema 2 decide ahora poner a prueba su programa de manera más detallada y es capaz de asignar a sus alumnos de manera aleatoria a un grupo que recibe una condición de control. También analiza los efectos del programa en los chicos y las chicas.
4. Un orientador quiere saber si existe una relación entre la autoestima de los alumnos de undécimo curso y la frecuencia de visitas al centro de orientación.
5. Un estudiante de doctorado está interesado en investigar las diferencias actitudinales entre alumnos con alto y bajo rendimiento que reciben diferentes tipos de retroalimentación por parte del profesor tras la realización de los exámenes. Para el estudio se han diseñado cuatro tipos de retroalimentación del profesor y hay ocho actitudes, tales como actitud hacia el profesor, hacia la asignatura, hacia la enseñanza, etc.



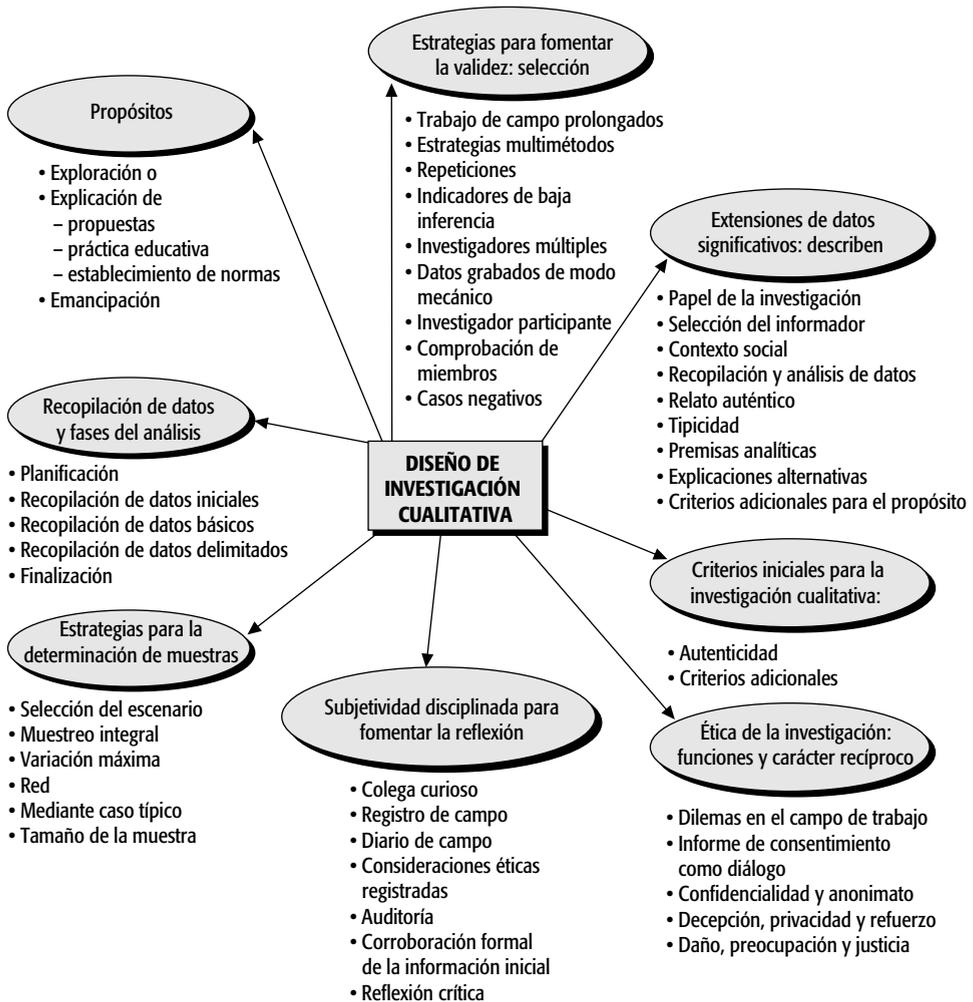
Diseños y métodos de investigación cualitativa

A menudo, los educadores se preguntan: ¿cómo diseñan sus estudios los investigadores cualitativos? ¿Cómo tratan estos investigadores asuntos de precisión como validez, subjetividad y empleo del estudio? ¿Cuándo se emplea la observación participativa o las entrevistas en profundidad? ¿Se deberían combinar diferentes estrategias de recogida de datos en un único estudio? ¿Cómo se lleva a cabo un análisis sistemático de los datos en la investigación cualitativa? ¿Cómo juzga un lector la credibilidad de los datos en un estudio cualitativo?

La investigación cualitativa aparecía clasificada en el capítulo 2 como una *investigación interactiva* o *investigación no interactiva*, denominada *investigación analítica*. La parte III trata de la investigación interactiva cualitativa y la parte IV presentará la investigación analítica.

En la parte III describimos el diseño y los criterios de investigación cualitativa (capítulo 10), las estrategias para la recogida de datos (capítulo 11) y los procesos y las técnicas de análisis de datos cualitativos (capítulo 12). Las estrategias de recogida de datos interactivos se utilizan, primordialmente, para el estudio de sucesos sociales actuales, ambientes y procesos sociales. El análisis de datos inductivos se elabora a partir de datos descriptivos de los significados de la gente, derivados de acontecimientos y procesos particulares o adscritos a ellos. La investigación cualitativa sugiere propuestas fundamentadas, aporta explicaciones para ampliar nuestro conocimiento de los fenómenos o promueve oportunidades para adoptar decisiones informadas para la acción social. La investigación cualitativa contribuye a la teoría, a la práctica educativa, a la elaboración de planes y a la concienciación social.

Introducción al diseño de investigación cualitativa



PALABRAS CLAVE

estrategias interactivas
investigación de campo
diseño de estudio de caso
diseño emergente
muestreo intencionado
selección del escenario
muestreo integral
muestreo de variación máxima
muestreo en cascada (de bola de nieve)

validez de los diseños cualitativos
subjetividad disciplinada
reflexibilidad
audibilidad
punto de vista
extensión de los hallazgos
autenticidad
tipicidad

La investigación interactiva cualitativa es el sondeo con el que los investigadores recopilan los datos en situaciones reales por interacción con personas seleccionadas en su propio entorno (investigación de campo). La investigación cualitativa describe y analiza las conductas sociales colectivas e individuales, las opiniones, los pensamientos y las percepciones. El investigador interpreta fenómenos según los valores que la gente le facilita. Los estudios cualitativos son importantes para la elaboración de la teoría, el desarrollo de las normas, el progreso de la práctica educativa, la explicación de temas sociales y el estímulo de conducta.

En este capítulo introduciremos el diseño cualitativo y relacionaremos el diseño con las cuestiones generales de investigación y las contribuciones potenciales del estudio. Explicaremos estrategias para la finalidad de la muestra y las fases de recopilación de datos y análisis. Se tratan varias maneras de abordar temas importantes: estrategias para fomentar la validez, técnicas para minimizar los sesgos de la investigación y el diseño de elementos para la extensión de los resultados. También trataremos la ética de la investigación en los estudios de campo y los patrones para juzgar la suficiencia de los diseños cualitativos. Indicaremos las distinciones entre los enfoques «tradicionales» y críticos de la investigación.

PROPÓSITOS, PREGUNTAS DE INVESTIGACIÓN Y DISEÑO DEL ESTUDIO DE CASO

Enfoque de investigación y orientación

Antes de que un investigador diseñe un estudio, ya posee una idea general del tema de la investigación y ha seleccionado un modo de investigar apropiado

para el estudio. Las diferencias generales entre investigación cualitativa y cuantitativa se resumieron ya en la primera parte. Los cinco enfoques de la investigación interactiva (modos de investigar) comentados en el capítulo 2 eran estudios etnográficos, fenomenológicos, estudio de caso, de teoría justificada y estudios críticos. Aquí, revisaremos la orientación de la investigación compartida entre estos modos de investigación porque proporciona una razón fundamental para muchas decisiones del diseño.

Suposiciones. La investigación cualitativa se basa en una filosofía constructivista que asume la realidad como una experiencia heterogénea, interactiva y socialmente compartida, interpretada por los individuos. Los investigadores cualitativos creen que la realidad es una «construcción social», es decir, los individuos o grupos deducen o atribuyen significados a entidades concretas, como acontecimientos, personas, procesos u objetos. Las personas elaboran construcciones para dar sentido a estas entidades y reorganizarlas según sus puntos de vista, sus percepciones y sistema de creencias. En otras palabras, las percepciones de las personas son lo que ellos consideran «real» para ellos y lo que dirige sus acciones, sus pensamientos y sus sentimientos.

Meta. La investigación cualitativa se ocupa, en primer lugar, de entender el fenómeno social desde la perspectiva de los participantes. El entendimiento se adquiere analizando los muchos contextos de los participantes y mediante la exposición de los significados de los participantes sobre estas situaciones y acontecimientos. Los significados de los participantes incluyen sus sentimientos, sus creencias, sus ideas, sus pensamientos y sus conductas. Por ejemplo, el aprendizaje de la lectura tiene lugar en el contexto de las escuelas, de las familias y en otras situaciones, e implica historias personales e interpersonales. Todo esto influye en el proceso de aprendizaje y el significado para un niño de lo que lee. Algunas investigaciones cualitativas pretenden hacer algo más que sólo entender los fenómenos y también crean teorías o refuerzos.

Estrategias con varios métodos. Los investigadores cualitativos estudian las perspectivas de los participantes con **estrategias interactivas**, (por ejemplo, la observación del participante, la observación directa, las entrevistas en profundidad, los instrumentos y técnicas suplementarias). Las estrategias de investigación son flexibles, pues emplean diversas combinaciones de técnicas para obtener datos válidos. La mayoría de los investigadores ajustan las decisiones sobre las estrategias de recopilación de datos durante el estudio. Las múltiples realidades se ven como tan complejas que nadie puede decidir *a priori* una singular metodología.

Función de la investigación. Los investigadores se «sumergen» en la situación y los fenómenos estudiados. Los investigadores asumen las funciones sociales interactivas en las que registran observaciones e interacciones con los participantes. La función del investigador varía desde la actitud más imparcial hasta una

disposición participativa y activa, dependiendo del enfoque de investigación seleccionado. Los escolares destacan la importancia de recopilar datos utilizando una «persona» preparada y hábil en lugar de una prueba única.

Sensibilidad del contexto. Otras características de la investigación cualitativa derivan de la creencia de que las acciones humanas están fuertemente influenciadas por la situación en la que tienen lugar. El estudio es una **investigación de campo**, es decir, el investigador recopila datos durante un espacio prolongado de tiempo en un escenario o a partir de individuos. Las generalizaciones contextuales están orientadas a ser utilizadas por los participantes, los lectores interesados o por otros investigadores en la posterior investigación con estudios de caso adicionales o diseños más estructurados.

Propósito y cuestiones de investigación

Históricamente, los investigadores cualitativos citaban dos propósitos principales de un estudio: *describir* y *explorar* y *describir* y *explicar*. Términos parecidos podrían ser «examinar» o «documentar», «entender», y «descubrir» o «generar». Muchos estudios cualitativos son descriptivos y exploratorios. Se añaden a la bibliografía por la elaboración de descripciones completas de situaciones complejas y señalando direcciones para la investigación futura. Otros estudios cualitativos son claramente explicativos. Muestran las relaciones entre los acontecimientos y los valores, normalmente tal y como los perciben los participantes. Estos estudios aumentan la comprensión de los fenómenos por parte del lector. Otras consideraciones del propósito dirigen la acción o el refuerzo que, a menudo, constituyen las metas últimas de los estudios críticos (investigación de acción participativa, crítica, feminista y posmodernista). Aunque los investigadores pueden requerir autorización y emprendiendo la acción como parte del propósito del estudio, en realidad, sólo pueden advertir cómo la investigación puede ofrecer oportunidades para la autorización informada (ver tabla 10.1).

Lógicamente, la pregunta general de la investigación está relacionada con el propósito. La mayoría de las preguntas cualitativas sobre tópicos analíticos implican el «cómo» y el «porqué» de unos fenómenos. Para tratar estos problemas el investigador recopila detalles descriptivos sobre el «quién», «qué», «dónde» y «cuándo» de los fenómenos. Estas observaciones incluyen el contenido y el caso orientado por el diseño. Con frecuencia, los investigadores cualitativos se centran en individuos, grupos, procesos y en organizaciones y sistemas.

Diseño de un estudio de caso

La investigación cualitativa utiliza un **diseño de estudio de caso** que significa que el análisis de los datos se centra en un fenómeno, seleccionado por el

TABLA 10.1: PROPÓSITO DE LA INVESTIGACIÓN Y PREGUNTAS ILUSTRATIVAS DE LA INVESTIGACIÓN

Propósito de la investigación	Preguntas ilustrativas de investigación
<p>Exploración descriptiva Para examinar fenómenos «nuevos» o poco conocidos.</p> <p>Para descubrir motivos de las intenciones de los participantes.</p> <p>Para desarrollar un concepto, modelo o hipótesis de forma detallada para la futura investigación.</p>	<p>¿Qué está pasando en esta situación social?</p> <p>¿Cuáles son las categorías y los motivos de las intenciones de los participantes?</p> <p>¿Cómo están relacionados estos patrones con las propuestas/ afirmaciones?</p>
<p>Explicación descriptiva Para describir y explicar los patrones relacionados con el fenómeno.</p> <p>Para identificar las relaciones que influyen sobre los fenómenos.</p>	<p>¿Qué acontecimientos, creencias, actitudes y/o normas influyen en este fenómeno?</p> <p>¿Cómo explican los participantes el fenómeno?</p>
<p>Emancipación Para crear oportunidades y la voluntad de iniciar la acción social.</p>	<p>¿Cómo explican y describen los participantes sus problemas y emprenden acciones positivas?</p>

Fuente: adaptado de Marshall y Rossman, 1999, p. 33.

investigador para entender independientemente del número de escenarios o de participantes en el estudio. El « uno » puede ser, por ejemplo, un administrador, un grupo de estudiantes, un programa, un proceso, la aplicación de una norma o de un concepto.

La investigación cualitativa requiere un plan para elegir escenarios y participantes, y para iniciar la recopilación de datos. El plan es un **diseño emergente**, en el que cada decisión de incrementar la investigación depende de la información previa. En realidad, el diseño emergente, puede parecer circular, como los procesos de determinación intencionada de propósitos, la recopilación de datos y el análisis parcial de datos son simultáneos e interactivos en lugar de discretos pasos secuenciales¹.

¹ Ver Bogdan y Biklen (1998), Giesne y Peshkin (1992), Marshall y Rossman (1999), Strauss y Corbin (1998), Stake (1995), La Compte y Preissle (1993) y Lincoln y Guba (2000).

Los investigadores cualitativos investigan grupos pequeños y distintos como todo el cuerpo docente de una escuela innovadora, todos los estudiantes de una clase seleccionada, el objetivo principal de un año académico o de una institución. Estos son estudios de escenarios únicos en los que hay un contexto socio-cultural natural y una interacción directa que incluyen a la persona o al grupo. Los investigadores cualitativos estudian, también, grupos de individuos que han tenido una experiencia similar pero no podrían estar interactuando entre ellos como las familias de niños que han sufrido abusos físicos.

A veces, lo que se trata es el contraste entre subunidades dentro de un programa, como los grupos demográficos (varón/ hembra o negro/ blanco) o grupos programados (graduados/ sujetos que han abandonado los estudios o los que trabajan bien/ mal), pero el propósito es entender el fenómeno «uno»: la entidad o el proceso.

Las subunidades son grupos de contraste que es probable que proporcionen información sobre la investigación². Estos grupos no son vistos como estadísticamente comparables y tampoco como exclusivos entre sí; a menudo son seleccionados para investigar la extensión o la diversidad del fenómeno. El investigador examina detalladamente el primer grupo y, a continuación, selecciona otro grupo para contrastar o colaborar con el primer grupo. La recopilación de datos sobre el segundo grupo requiere menos tiempo porque el investigador ya ha limitado los parámetros del estudio. Los grupos no se establecen de manera simultánea a causa de la complejidad y la dificultad de organizar la recopilación de datos.

En el diseño del caso de estudio del extracto 10.1 el investigador selecciona a una profesora de ciencias, Sarah, para un estudio de colaboración. El propósito era entender o «dar sentido» a la actividad de Sarah y a las actividades de sus alumnos, de sus compañeros profesores y de los directores de la escuela en los «términos de lo que ella cree y cómo construyó los varios contextos en los que enseñó» (Tobin y La Master, 1997, p. 227).

EXTRACTO 10.1 **EL DISEÑO DEL ESTUDIO DE CASO**

La metodología de la investigación empleada en el estudio era interpretativa y se esforzó por dar sentido a la cultura de las aulas en las que Sarah impartía sus clases, a partir de las actividades de los participantes (por ejemplo, Sarah, los alumnos, los compañeros de los profesores y los directores de la escuela). Una argumentación del estudio

² Bogdan y Biklen (1998) reclaman un examen de subunidades en un estudio de multicaso. Un estudio de multicaso no es un estudio sobre muchos lugares que emplea un enfoque inductivo analítico, con frecuencia con muchos investigadores.

fue para dar sentido a las actividades de Sarah mediante la descripción de sus comportamientos y el esfuerzo por darles sentido en términos de lo que creía y cómo construyó los varios contextos en los que enseñaba. Nuestra opinión sobre la investigación de colaboración requería que escucháramos las voces de Sarah y de otros participantes en el estudio y también designar una clara «firma» (Cladinin y Connelly, 1994) a lo que aprendimos en términos de interpretación que reflejaban las perspectivas de ambos autores.

Significado y justificación

Planear un diseño de estudio de caso implica, no sólo, seleccionar la cuestión general de la investigación, sino también incorporar componentes del diseño que se añaden a las posibles contribuciones y al significado del estudio. La investigación cualitativa puede ser diseñada para contribuir a la teoría, la práctica, las normas y los temas y acciones sociales. Describiremos a continuación cada uno de estos apartados junto con otras justificaciones.

Contribuciones a la teoría. El diseño de estudio de caso resulta apropiado para la investigación exploratoria y orientada al descubrimiento. Los *estudios exploratorios* que examinan un tema sobre el que ha habido poca investigación previa, se diseñan para alcanzar una investigación más detallada. El propósito es elaborar un concepto, desarrollar un modelo con sus subcomponentes relacionados o sugerir propuestas. Algunos estudios facilitan el entendimiento de un concepto abstracto, como la gestión basada en la escuela, a partir de la experiencia social de los participantes. Otros estudios relacionan las percepciones de los participantes sobre la ciencia social y sugieren propuestas sobre las personas en general, en lugar de relacionar los datos con un concepto educativo. Los conceptos, modelos o hipótesis son «teorías argumentadas» porque estos conceptos abstractos se elaboran a partir de observaciones en lugar de deducirse a partir de teorías previas.

Contribuciones a la práctica. Los estudios cualitativos pueden aportar una descripción detallada y un análisis de una práctica particular, de un proceso o de un acontecimiento. Algunos estudios informan sobre los sucesos mientras que otros contribuyen aumentando el propio entendimiento de los participantes acerca de la práctica para mejorarla. Una serie de estudios cualitativos con enfoques parecidos sobre la investigación, orientados de *forma independiente* por diferentes investigadores en diversas situaciones *en un espacio de tiempo de años*, pueden contribuir al conocimiento educativo a través del predominio de la evidencia acumulada. Áreas específicas de la educación para las que los diseños cuantitativos resultaban inadecuados han comenzado a acumular evidencia con los estudios de caso. Un ejemplo es una serie de estudios de caso sobre la función de un director de escuela elemental que indica las complejidades de este puesto administrativo. Hasta

que las dificultades del diseño cualitativo puedan resolverse, el conocimiento basado en un predominio de la evidencia no puede ser fácilmente ignorado.

Contribuciones a las normas. La investigación cualitativa que emplea un diseño de estudio de caso contribuye, también, a la formulación de normas, su cumplimiento y su modificación. Algunos estudios se centran en los informales procesos de la formulación o el cumplimiento de normas en diferentes situaciones con diversos valores culturales para explicar los resultados de las normas públicas. La investigación cualitativa puede analizar las percepciones de los influyentes sobre un asunto económico o político de la colectividad, las actitudes de los que elaboran las normas según una fórmula propuesta y las opiniones de aquéllos que cumplen las normas. Con frecuencia, estos estudios identifican las consecuencias que inspiran la necesidad de modificar estatutos o regulaciones y ayudar a los que elaboran las normas a anticipar futuros temas.

Contribuciones a temas sociales y acción. A menudo, los estudios críticos están destinados a la revisión y la transformación, el desgaste de la ignorancia y el refuerzo. Algunos estudios se centran en la experiencia vivida por grupos raciales y étnicos, clases sociales, y los papeles de género. Los investigadores sitúan categorías como «raza», «grupo étnico», «clase social», «homosexual» y «hembra» en un contexto social más general para criticar sus aspectos ideológicos y los intereses políticos/económicos que se benefician de la situación. Con frecuencia, los estudios expresan la «cultura del silencio» de varios grupos; otros describen formas de resistencia y organización de grupos que desarrollan sus propios valores como una fuerza para la cohesión y la supervivencia en la cultura dominante.

Otras justificaciones. La investigación cualitativa y el diseño de estudio de caso pueden justificarse por la vialidad de los resultados relacionados con la obtención de datos válidos. La investigación cualitativa se hace característica cuando la naturaleza de la situación o los individuos no permiten utilizar una prueba. Las estrategias cualitativas son apropiadas con personas que están extremadamente ocupadas o personas que son expresivas de manera no verbal o utilizan un segundo idioma. La investigación cualitativa se realiza, también, cuando el tema es confidencial o genera controversia dentro de una institución y se asegura sólo la documentación mínima. En algunas situaciones, un estudio experimental no puede hacerse por razones prácticas o éticas. No hay un número suficiente de participantes para reconocer los requisitos estadísticos ni para establecer un grupo de comparación.

MUESTREO INTENCIONADO

El muestreo intencionado, al contrario que la determinación de probabilidad, consiste en «seleccionar casos con abundante información para estudios detallados» (Patton, 1990, p.169) cuando alguien pretende entender algo sobre estos casos sin necesitar o desear generalizar sobre cada uno de los casos. El muestreo

intencionado se realiza para aumentar la utilidad de la información obtenida a partir de pequeños modelos. El muestreo intencionado requiere que la información se obtenga sobre variaciones entre las subunidades antes de que se elija el modelo. Entonces, el investigador busca para la «información abundante» informadores clave, grupos, escenarios o acontecimientos para estudiarlos. En otras palabras, estos modelos se escogen porque es probable que sean inteligibles e informativos sobre los fenómenos que el investigador está investigando.

El poder y la lógica del muestreo intencionado consisten en que, con pocos casos estudiados en profundidad, se obtienen muchas aclaraciones sobre el tema; mientras, la lógica de la determinación de muestras probables depende de la selección de un modelo representativo estadísticamente o aleatorio para la generalización a una población mas numerosa. Los procedimientos de determinación de muestras de probabilidad tales como un modelo simple aleatorio o un modelo estratificado pueden resultar inapropiados cuando (1) el propósito no es la generalizabilidad de la información (2) sólo son relevantes para el problema de la investigación una dos subunidades de una población; (3) los investigadores no tienen acceso a todo el grupo a partir del cual desean establecer el modelo; o (4) resulta imposible la determinación estadística de modelos a causa de razones logísticas y éticas.

Los tipos de muestreo intencionado incluyen la selección del escenario, el muestreo comprensible, el muestreo de variación máxima, el muestreo en cascada y el muestreo mediante tipo de caso (ver tabla 10.2).

Selección del escenario

La **selección del escenario** mediante la cual se designa un espacio para localizar personas implicadas en un acontecimiento concreto, es preferible cuando la investigación se centra en microprocesos complejos. Resulta esencial una clara definición de los criterios para la selección del escenario. Los criterios están relacionados y son apropiados para el tema de la investigación y para el propósito. Por ejemplo, si se utiliza la cuestión inicial para describir y analizar la toma de decisiones de los profesores sobre actividades de aprendizaje, o perspectivas y estrategias de los alumnos con respecto a la organización de las clases, o el concepto de profesión por parte de los profesores de la etapa elemental, entonces, el escenario seleccionado debería hacer posible que estos puntos de vista o acciones estuvieran presentes y pudieran estudiarse.

Muestreo comprensible

El **muestreo comprensible**, en el que cada participante, grupo, situación, acontecimiento u otra información relevante se examina, es la estrategia de determinación de muestras preferida. Cada subunidad se maneja por categorías y lo

TABLA 10.2: ESTRATEGIAS PARA EL MUESTREO INTENCIONADO

Estrategia del modelo	Descripción
Selección del escenario	Seleccionar el espacio donde se espera que tengan lugar los acontecimientos concretos.
Muestreo comprensible	Elegir un grupo entero a partir de criterios.
Muestreo de variación máxima	Seleccionar para obtener las diferencias máximas de percepciones sobre un asunto entre informadores con abundante información o grupo.
Muestreo en cascada	Cada persona o grupo sucesivo está designado por una persona importante como apropiados para un perfil o atributo.
Muestreo mediante el tipo de vaso	
Caso extremo	Elegir casos extremos tras conocer el caso típico o aproximado, por ejemplo, sucesos destacados, acontecimientos críticos.
Caso exagerado	Seleccionar casos que resultan exagerados pero no extremados, por ejemplo, bajos promedios de alumnos.
Caso típico	Conocer las principales características de un grupo o de una muestra, por ejemplo, la selección de un director de un instituto tradicional eliminaría a las mujeres, a las personas demasiado jóvenes o demasiado viejas, a los varones solteros.
Caso único	Elegir el caso raro o poco frecuente de alguna dimensión o acontecimiento, por ejemplo, la aplicación de un nuevo decreto de normas federales.
Caso destacado	Obtener la recomendación de los expertos para seleccionar los mejores, por ejemplo, el director designa los profesores «competentes» y funcionarios estatales para identificar escuelas «eficaces».
Caso crítico	Identificar el caso que puede ilustrar algún fenómeno de manera dramática, por ejemplo, el caso del «examen real» o el caso «ideal».
Concepto/base teórica	Seleccionar mediante abundante información personas o situaciones conocidas para experimentar el concepto o intentar cumplir el concepto/teoría, por ejemplo, escuela que se administra en base a su emplazamiento, <i>burnout</i> del profesor.
Combinación de estrategias de muestreo intencionado	Elegir diversos modelos de estrategias, tantos como necesitados o requeridos por los objetivos, especialmente en los estudios a gran escala y estudios de procesos de larga duración.

suficientemente diferenciadas para que no se pierdan posibles variaciones. Por ejemplo, un estudio de una corriente dominante de niños autistas en una escuela requeriría, probablemente, la observación de todos los niños autistas. Supongamos que un estudio sobre alumnos internos de instituto en un programa de aprendizaje externo se realizaba en 35 lugares diferentes. Cada lugar de trabajo era tan diferente –un discurso clínico de hospital, un periódico local, un sindicato, dos oficinas legislativas, un refugio de animales y otros– la selección global sería necesaria. Los investigadores utilizan otras estrategias para la determinación de muestras porque los grupos, frecuentemente, no son lo bastante pequeños y los recursos raramente son suficientes.

Muestreo de variación máxima

El **muestreo de variación máxima** o la selección por cuotas, es una estrategia para aclarar diferentes aspectos sobre la cuestión el problema de la investigación. Por ejemplo, un investigador podría dividir una población de profesores de escuela elemental según el número de años de servicio en tres categorías y seleccionar informadores clave en cada categoría para investigar el desarrollo profesional. Éste no es un modelo representativo porque el investigador cualitativo utiliza tan sólo esta estrategia para describir, de forma detallada, los diferentes significados de desarrollo profesional de profesor entre individuos con diferentes años de servicio. Ver el extracto 10.2 para una combinación de estrategias de muestreo intencionado.

EXTRACTO 10.2

MUESTREO INTENCIONADO: NOMINACIÓN Y VARIACIÓN MÁXIMA

Los participantes en este estudio eran padres de alumnos que se graduaron en un programa para personas con discapacidades cognitivas entre 1989 y 1993. El modelo incluía padres cuyos hijos asistían a tres de los cuatro institutos de la ciudad y que tenían discapacidades cognitivas leves, moderadas y graves. Al emplear el «muestreo intencionado»... se seleccionaba a los participantes para que representaran tres escuelas, la presencia de las desventajas socioeconómicas, el grado de gravedad de la discapacidad de los niños y el nivel actual de los niños en la escuela. Se desarrolló un esbozo del modelo... con la asistencia de los educadores especiales del distrito escolar. Los padres sin número de teléfono resultaban eliminados del esbozo del modelo. Veinticuatro nombres... fueron seleccionados, contactados por los investigadores, y se les pidió que participaran en el estudio. El número de participantes estaba basado en las consideraciones de tiempo y viabilidad. De las veinticuatro familias contactadas, 19 aceptaron participar.

Fuente: Hanley-Maxwell, C., Whitney-Thomas, J. y Pogooff, S.M (1995), «The second shock: A qualitative study of parents' perspectives and needs during their child's transition from school to adult life» (p. 5), *JASH Journal of the Association for Persons with Severe Handicaps*, 20(1), 3-15.

Muestreo en cascada

El **muestreo en cascada**, también denominado **de bola de nieve**, es una estrategia en la que cada sucesivo participante o grupo es designado por el participante o grupo precedente. El investigador elabora un perfil de las particularidades o de una característica concreta buscada y solicita a los participantes que sugieran otras personas que se ajusten al perfil o posean esta particularidad. Esta estrategia puede utilizarse en situaciones en las que los individuos buscados no componen un grupo naturalmente limitado sino que están repartidos entre otras poblaciones. El muestreo en cascada se utiliza con frecuencia para los estudios de entrevistas en profundidad en lugar de una investigación de observación del participante.

Muestreo mediante tipo de caso

Cuando un estudio requiere un examen de un tipo de caso particular se utilizan otras estrategias para la determinación de muestras. Recordemos que «caso» se refiere a un análisis profundo de un fenómeno y no al número de personas tomadas como muestra. Los ejemplos de determinación de muestras mediante el tipo de caso son: caso extremo, caso exagerado, caso típico, caso único, caso destacado, caso crítico y determinación del modelo concepto/ base teórica. Cada una de estas estrategias para el muestreo aparece definida en la tabla 10.2. Un investigador podría elegir combinaciones de tipos de casos como necesitados o deseados, especialmente en estudios a gran escala y en estudios de procesos prolongados.

Las estrategias para la determinación de los modelos intencionados empleados en un estudio se identifican a partir de la información principal y se presentan en el estudio para mejorar la calidad de los datos. Además, las personas o grupos que participaron realmente en el estudio se presentan de manera que se proteja la confidencialidad de los datos. Los investigadores históricos y legales especifican las recopilaciones utilizadas de los archivos públicos y privados y, con frecuencia, se refieren a cada documento o caso breve mediante pies de página explicatorios. De este modo, los investigadores utilizan técnicas no interactivas para estudiar si el pasado reduce las amenazas de validez del diseño.

El tamaño de la muestra

Los investigadores cualitativos consideran los procesos de la determinación de la muestra como *dinámicos*, *adecuados* y *básicos* más que como parámetros de población estáticos o previos. Mientras existen normas estadísticas para el tamaño de la muestra de probabilidad, para el tamaño de los modelos intencionados sólo existen líneas de guía. Los modelos intencionados pueden estar ordenados desde $n = 1$ hasta $n = 40$ o más. Tradicionalmente, un tamaño de muestra

cualitativa parece pequeño en comparación con el tamaño del modelo necesario de representatividad para generalizar a una población más numerosa.

La lógica del tamaño de la muestra está relacionada con el propósito del estudio, el problema de la investigación, la técnica de recopilación de los datos principales y la disponibilidad de los casos con abundante información. La comprensión generada a partir de la investigación cualitativa depende más de *la abundancia de información acerca de los casos y de las capacidades analíticas del investigador* que del tamaño de la muestra.

Los investigadores utilizan las líneas de guía siguientes para determinar el tamaño de la muestra.

1. ¿Cuál es el propósito del estudio? Un estudio de caso que es exploratorio o descriptivo puede no necesitar tantos casos como un estudio que se autolimita y busca una explicación/descripción. Además, normalmente un estudio fenomenológico tiene menos informadores que los que precisa una teoría argumentada para generar conceptos densos.
2. ¿Cuál es la orientación del estudio? Los estudios centrados en el proceso dependen de la duración normal del proceso y, a menudo, tienen menos participantes; mientras que un estudio mediante entrevista a informadores seleccionados depende del acceso a los informadores.
3. ¿Cuál es la estrategia de recopilación de datos previos? A menudo, los investigadores cualitativos discuten el número de días en el campo, si se dedican a la observación o a la entrevista. Algunos estudios podrían tener una muestra de tamaño pequeño, pero el investigador vuelve constantemente a la situación o a los mismos informadores en busca de la confirmación de los datos.
4. ¿Cuál es la disponibilidad de los informadores? Algunos casos son raros y difíciles de localizar; otros casos son relativamente fáciles de identificar y de localizar.
5. ¿La información se está haciendo redundante? ¿Podrían añadirse más informadores o volver al campo para recoger alguna nueva comprensión?
6. Los investigadores se someten el tamaño de la muestra obtenida para averiguar la revisión o el juicio global. La mayoría de los investigadores cualitativos proponen un tamaño «mínimo» de muestra y, a continuación, siguen ampliando la muestra a medida que progresa el estudio.

FASES DE RECOPIACIÓN DE DATOS Y ESTRATEGIAS DE ANÁLISIS

Las fases cualitativas de la recopilación de datos y los análisis son procesos de investigación interactivos que tienen lugar en ciclos superpuestos. Estos no se

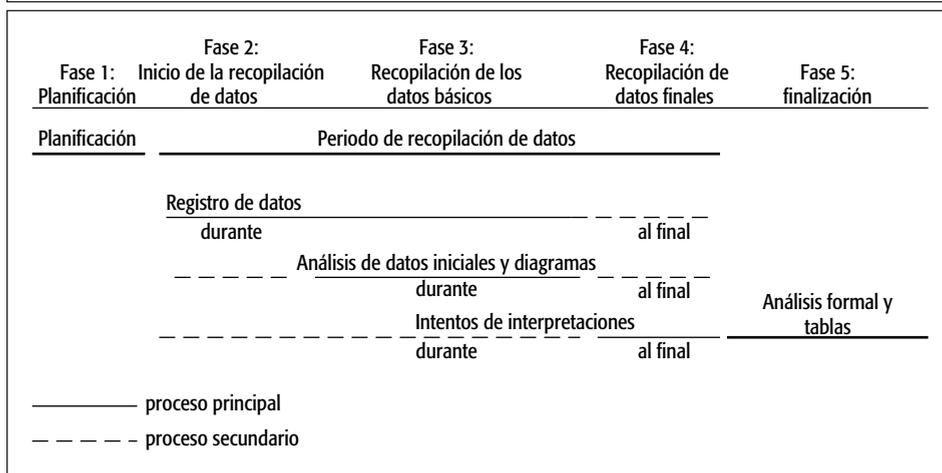
llaman procedimientos sino «recopilación de datos y estrategias de análisis», técnicas que son flexibles y dependientes de cada estrategia principal y de los datos obtenidos a partir de esta estrategia. Las fases de la investigación son relativamente parecidas para los diferentes modos de averiguación cualitativa. La figura 10.1 representa cinco fases de investigación: la fase 1 es la planificación; las fases 2, 3, y 4 son el inicio, la base y la recopilación de datos; y la fase 5 es la finalización. Las cinco fases de la investigación, explicadas a continuación, demuestran los procesos interactivos de determinación de muestras y de selección, de registro de datos, de análisis y de exposición, y de tentativas de interpretación durante el periodo de la recopilación de datos.

Fase 1: planificación. Para planear un estudio cualitativo, los investigadores analizan la presentación del problema y las preguntas de la investigación, que orientan los esfuerzos de la recopilación de datos. A continuación, describen el tipo de situación, escenarios o entrevistas que, lógicamente, parece que ofrecen información sobre el problema. Esta descripción se convierte en la línea guía para el muestreo intencionado y para la selección. En la fase 1, un investigador localiza y obtiene el permiso para utilizar el escenario o una red de personas.

Fase 2: inicio de la recopilación de datos. Esta fase incluye los primeros días en el campo durante los que un investigador elabora un informe de relaciones verdaderas y recíprocas con los individuos y los grupos para observarlos (Wax, 1971). En primer lugar, el investigador obtiene los datos para orientarse en el campo y conseguir un sentido de «totalidad» de la situación para el muestreo intencionado.

En la fase 2 el entrevistador interroga a las primeras personas de la red, y comienza la técnica del muestreo con efecto «bola de nieve». Al mismo tiempo,

FIGURA 10.1: FASES DE LA INVESTIGACIÓN CUALITATIVA



el entrevistador perfecciona la entrevista y los procedimientos de registro. Se realizan ajustes en las técnicas del entrevistador para establecer un informe y en el orden y presentación de preguntas/consideraciones durante las entrevistas.

Al principio del estudio, el investigador cualitativo, independientemente de la técnica de la recopilación de datos, desarrolla una manera de organizar, de codificar y de reunir datos recopilados para el análisis formal de los datos (fase 5). Las transcripciones de notas de campo o de las entrevistas y de los registros de los documentos del investigador contienen, a menudo, miles de páginas impresas.

Fase 3: recopilación de datos básicos. El investigador no se enreda durante mucho tiempo con los ajustes sobre las novedades de la situación del campo o sobre las peculiaridades de las entrevistas.

En la fase de recopilación de datos básicos, el que realiza las preguntas empieza a «oír», «ver» y «leer» lo que está sucediendo, en vez de sólo escuchar, mirar alrededor o escanear documentos. El investigador sigue seleccionando estrategias para la recopilación de datos y de los informadores.

El intento de análisis de los datos comienza cuando el investigador procesa mentalmente muchas ideas y hechos durante la recopilación de los datos. Las concepciones iniciales sobre el trabajo y las descripciones se transforman y resumen, con frecuencia, mediante la elaboración de diagramas integradores del trabajo. Cuando aparecen los patrones iniciales, el investigador identifica las ideas y los hechos que necesitan una corroboración en la fase final.

Fase 4: recopilación de datos finales. La recopilación de datos se completa cuando el investigador «abandona el campo» o conduce la última entrevista. En la investigación cualitativa, no existen datos *a priori* hasta el final de la recopilación de datos como podría suceder en estudios cuantitativos, cuando un cierto porcentaje de cuestionarios completados se devuelve. La recopilación de datos finales está relacionada con el problema de la investigación y con la profundidad y la riqueza de los datos recopilados. En la fase 4, el investigador presta más atención a las posibles interpretaciones y a la verificación de la información inicial con los informadores clave, las entrevistas restantes o los documentos. Como advierten Glaser y Strauss (1967):

La continua búsqueda de recopilación de datos y su análisis tiene un apoyo directo en cómo se completa la investigación... [El investigador] cree en su propio conocimiento... no a causa de un juicio arbitrario sino porque se ha esforzado especialmente en descubrir lo que piensa que podría saber, cada paso del camino desde el principio de su investigación hasta sus conclusiones... Ha estado viviendo durante muchos meses con análisis parciales, probándolos en cada uno de los pasos, hasta que ha construido su teoría (interpretaciones). Es más, si ha participado (o destacado) en la vida social de este sujeto, entonces ha estado viviendo mediante sus análisis, probándolos no sólo mediante la observación y la entrevista (y la lectura) sino también mediante la vida diaria (p. 224-225).

El investigador siente que una mayor recopilación de datos puede aportar más datos pero no más descubrimientos relevantes al problema de la investigación.

Fase 5: finalización. La finalización de la fase de la recopilación activa de datos da lugar al análisis formal de datos y a la elaboración de formas significativas de presentar los datos. El análisis de los datos empieza con una construcción de «los hechos» como se encontraron en los datos registrados por el investigador. El investigador reconstruye diagramas iniciales, gráficos temporales, listas de frecuencia, tablas del proceso y otros para sintetizar un sentido global de la «totalidad», la relación de las partes con el todo. El investigador plantea una serie de preguntas sobre los datos registrados mientras «introduce» lentamente temas, interpretaciones y/o propuestas. El análisis de los datos y las tablas resultan esenciales antes de que el investigador pueda hacer interpretaciones.

VALIDEZ DE LOS DISEÑOS CUALITATIVOS

La validez se refiere al grado en que las explicaciones de los fenómenos determinan las realidades del mundo. La validez en la investigación cuantitativa incluye tanto la interna (deducciones causales) como la externa (generalizabilidad) y cuestiones de objetividad y de credibilidad. La investigación cualitativa emplea diferentes suposiciones, diseños y métodos para desarrollar los conocimientos. Aunque existe un acuerdo general en el empleo de conceptos diferentes para los patrones de investigación cualitativa, los desacuerdos aparecen en los nombres para los conceptos específicos y la importancia relativa de cada criterio. Utilizamos términos generales y reconocidos, es decir, la validez, la subjetividad disciplinada y la extensión de los descubrimientos como los criterios más frecuentes para la investigación cualitativa.

La validez del diseño cualitativo

La validez se encarga de estas preguntas: ¿observan, realmente, los investigadores lo que ellos creen que observan? ¿Oyen, realmente, los investigadores los significados que creen que oyen? En otras palabras, **la validez de los diseños cualitativos** es el grado en el que las interpretaciones y los conceptos poseen *significados recíprocos* entre los participantes y el investigador. El investigador y los participantes están de acuerdo en la descripción o la constitución de los acontecimientos, especialmente en los significados de estos acontecimientos.

Estrategias para mejorar la validez

En la investigación cualitativa, las exigencias de validez se limitan a las técnicas de recopilación de datos y de análisis. Los investigadores utilizan una

combinación de cualquiera de las diez estrategias posibles para mejorar la validez: campo prolongado de trabajo, estrategias con varios métodos, lenguaje repetitivo del participante, indicadores de nivel bajo de deducción, investigadores múltiples, datos registrados mecánicamente, investigador participante, comprobación de miembros, revisión de participantes y casos negativos (ver tabla 10.3)

TABLA 10.3: ESTRATEGIAS PARA MEJORAR LA VALIDEZ DEL DISEÑO: ESTRATEGIAS DE RECOPIACIÓN DE DATOS PARA AUMENTAR LA CONFORMIDAD SOBRE LA DESCRIPCIÓN O LA CONSTITUCIÓN DE LOS FENÓMENOS ENTRE INVESTIGADORES Y PARTICIPANTES

Estrategia	Descripción
Campo de trabajo prolongado y continuo	Permite el análisis interno de los datos y la corroboración para asegurar la concordancia entre información y la realidad del participante.
Estrategias con varios métodos	Permite la triangulación en la recopilación de datos y el análisis de datos.
Lenguaje del participante; recuentos de repeticiones	Obtiene consideraciones literales de los participantes y citas de los documentos.
Indicadores de baja inferencia	Registra descripciones detalladas y casi literales de las personas y situaciones.
Investigadores múltiples	Coincide en los datos descriptivos recogidos por un equipo de investigación.
Datos registrados de manera mecánica	Empleo de grabadoras, fotografías y aparatos de video.
Investigador participante	Utilización de impresiones de los participantes registradas en diarios o recuerdos anecdóticos para la corroboración.
Comprobación de miembros	Comprobación informal con los participantes para precisar durante la recopilación de datos; realizada, con frecuencia, mediante estudios de observación de los participantes.
Revisión de participantes	Pregunta a cada uno de los participante para revisar los resúmenes de todas las entrevistas para precisar la representación; realizada, con frecuencia, mediante estudios de entrevistas.
Casos negativos o datos discrepantes	Búsqueda activa para, registrar, analizar, y presentar casos negativos o datos discrepantes que son una excepción en los modelos o que modifican los modelos encontrados en los datos.

Tradicionalmente, los investigadores cualitativos utilizan tantas estrategias como les resulta posible para asegurar un diseño válido. Las estrategias esenciales son prolongados campos de trabajo, estrategias múltiples, recuentos de repeticiones, indicadores de baja inferencia y búsqueda de casos negativos. Se añaden otras estrategias por ser apropiadas para mantener la cantidad mínima de interferencia mientras aumenta la calidad de los datos. La validez de los diseños cualitativos implica también cuestiones de ética y de viabilidad a causa de la variedad de diseños, preguntas de investigación y situaciones.

Campo de trabajo prolongado y continuo. La observación del participante y las entrevistas en profundidad se realizan en situaciones normales para reflejar la realidad de la experiencia de vida de forma más exacta a como se hace artificialmente en las situaciones de laboratorio. Los documentos históricos y legales son recuerdos de acontecimientos que ocurrieron en situaciones habituales. La extensión de la recopilación de datos ofrece oportunidades para los análisis internos de datos, las comparaciones preliminares y la corroboración para afinar las ideas y asegurar la concordancia entre las categorías basadas en la investigación y la realidad del participante.

Estrategias con varios métodos. La mayoría de los investigadores interactivos emplean varias técnicas de recopilación de datos en un estudio, pero, normalmente, seleccionan una como método central: la observación del participante o las entrevistas en profundidad. En cierta medida la observación del participante, la observación abierta, la entrevista y los documentos son un entrelazado en la web de todas las técnicas interactivas. Cómo varían cada una de estas estrategias utilizadas con el estudio. Además, pueden emplearse la recopilación de artefactos y las técnicas suplementarias (ver capítulo 11). Las estrategias con varios métodos permiten la *triangulación* de los datos a lo largo de las técnicas de averiguación. Estrategias diferentes pueden aportar conocimientos diferentes sobre el asunto que nos interesa y aumentar la credibilidad de la información. En este sentido general, *triangulación* puede también referirse al empleo de investigadores múltiples, de teorías múltiples o de perspectivas para interpretar los datos; las fuentes de datos múltiples para corroborar los pormenores (ver capítulo 12) y las disciplinas múltiples para ampliar el conocimiento personal del método y el fenómeno de interés (Janesick, 1998). Ver el extracto 10.3 como ejemplo de selección de estrategias para mejorar la validez del diseño.

El lenguaje del participante y el recuento de la repetición. Las entrevistas de los informadores, en el lenguaje de los participantes, resultan menos abstractas que muchas pruebas utilizadas en otros diseños. Para conseguir los nombres de los participantes para los acontecimientos y los objetos en sus experiencias diarias, se evita el empleo del lenguaje de ciencias sociales más abstracto. Los investigadores son también sensibles a los «traductores culturales», es decir,

informadores que traducen sus palabras en términos de ciencias sociales o del lenguaje social de clase. Por ejemplo, cuando se preguntó a los vagabundos: «¿dónde está tu casa?», «¿cuál es tu dirección?» traducirían la pregunta refiriéndose a una residencia típica de clase media y responderían «no tengo casa». Sin embargo, un investigador de campo encontró que los llamados «sin techo» por los científicos sociales, poseían «tiendas» que para ellos funcionaban como los «hogares» para las otras clases sociales (Spradley, 1979).

Los recuentos de las repeticiones de las conversaciones, las transcripciones y las citas directas a partir de documentos son muy valorados como datos. Los investigadores presentan en sus estudios extensas citas directas a partir de los datos para ilustrar los significados de los participantes.

Indicadores de nivel bajo de deducción. Las descripciones concretas y precisas de las notas del campo y las elaboraciones de la entrevista son las características distintivas de la investigación cualitativa y el método principal para identificar los patrones en los datos. «Baja inferencia» significa que las descripciones son casi literales y que los términos importantes son los que los participantes utilizan y entienden. Las descripciones de baja inferencia contrastan con el lenguaje abstracto del investigador.

EXTRACTO 10.3

MEJORA DE LA VALIDEZ

El estudio empleó un diseño interpretativo (Erickson, 1986) que seguía un ciclo hermenéutico en el que lo que se aprendía se obtenía a partir de lo que ya se sabía, la lectura de la bibliografía, la experiencia en el campo y los continuos esbozos de los datos, análisis e interpretaciones. Se adoptó un número de procedimientos, como la triangulación, para asegurar que el estudio tenía lo que Guba y Lincoln (1989) llamaban confirmabilidad (validez). La triangulación, que implica el empleo de numerosas fuentes de datos, maximizó la probabilidad de que las afirmaciones iniciales coincidieran con una variedad de datos. Como estuvimos en el campo durante un prolongado espacio de tiempo, se minimizó la tendencia de los participantes en el estudio a exhibir comportamientos artificiales en beneficio de los investigadores. Además, los investigadores eran capaces de distinguir si los comportamientos mostrados eran típicos o atípicos.

Fuente: Tobin, K. y LaMaster, S. U. (1995), «Relationships between metaphors, beliefs, and actions in a context of science currículo change» (p. 227-228), *Journal of Research in Science Teaching*, 32(3), 225-242.

Investigadores múltiples. El uso de investigadores múltiples es un método para mejorar la validez. El empleo de más de un investigador se trata de diferentes maneras: (1) entrenamiento inicial extensivo y discusión durante el trabajo de campo para llegar a acuerdos sobre los significados, (2) observaciones a corto plazo para la confirmación en escenarios diferentes y (3) con más frecuencia, un acuerdo

mediante el cual cada observador de campo es independientemente responsable de un lugar de la investigación y se encuentra periódicamente con el equipo para compartir las ideas y estrategias iniciales. Sin embargo, la investigación cualitativa basada en un enfoque amplio de grupo en equipo, se da con escasa frecuencia; la mayoría de los estudios cuenta sólo con dos investigadores como equipo.

Datos registrados de manera mecánica. Las grabadoras, fotografías y cámaras de video pueden mejorar la validez aportando una grabación precisa y relativamente completa. Sin embargo, para que los datos sean útiles, se deben observar los aspectos ambientales que afectaron a la grabación de los datos y que podrían facilitar la interpretación de los datos. Los aspectos ambientales pueden derivarse del equipamiento, de los ángulos o de la distancia de la grabadora, de la consecuencia del empleo del equipamiento técnico en el escenario social y sobre los participantes y otros.

El investigador participante. Muchos investigadores obtienen la ayuda de un informador para corroborar lo que han estado observando y registrando, las interpretaciones de los significados de los participantes y las explicaciones de los procesos generales. Los participantes podrían escribir diarios o realizar grabaciones anecdóticas para compartirlas con el investigador.

Comprobación de los miembros. Con frecuencia, los investigadores que establecen una residencia de campo confirman las observaciones y los significados de los participantes a través de conversaciones casuales con individuos en situaciones informales. La comprobación de los miembros puede también realizarse mediante una entrevista en la que los temas son replanteados y probados para obtener significados más completos y sutiles.

Revisión de los participantes. Los investigadores que entrevistan a cada persona o conducen una serie de entrevistas con la misma persona podrían solicitar a la persona que revisara la transcripción o el resumen de los datos obtenidos de él o ella. Se pide al participante que modifique cualquier información o interpretación de los datos de los entrevistados. A continuación, se analizan los datos obtenidos de cada entrevista para una integración de información comprensiva.

Casos negativos y/o discrepantes. Los investigadores buscan activamente, casos negativos registrados, analizados y presentados y, también, datos discrepantes. Un caso negativo es una situación, una situación social o una opinión de un participante que contradice el patrón inicial de significados. Los datos discrepantes presentan una variante al patrón inicial. Por ejemplo, un etnógrafo escolar podría encontrar que las habilidades interaccionan libremente entre ellas en seis situaciones. No se encontraron casos (situaciones) negativos. Sin embargo, los datos discrepantes, podrían sugerir que las interacciones entre las habilidades son «libres» en cinco situaciones y sólo «semilibres» en la sexta situación, dependiendo de quien esté presente.

SUBJETIVIDAD DISCIPLINADA EN LA INVESTIGACIÓN CUALITATIVA

La **subjetividad disciplinada** (Erickson, 1973) es el cuestionario personal del investigador y el uso de la empatía personal experimental en la recopilación de datos. La subjetividad facilita la relación entre los participantes. La autorreflexión sobre las propias opiniones, las preferencias personales y las predisposiciones teóricas pueden resultar positivas para cualquier tipo de investigación. La subjetividad disciplinada recuerda a muchos investigadores que quien investiga forma parte de la situación, del contexto y de los fenómenos sociales que él o ella pretende comprender.

La reflexión es un concepto relacionado y amplio que incluye un examen riguroso de los compromisos personales y teóricos para ver cómo sirven como recursos para seleccionar una de los muchos enfoques cualitativos, esbozando el problema de la investigación, generando datos particulares, las maneras de relacionarse con los participantes, y para desarrollar interpretaciones específicas (Altheide y Johnson, 1998; LeCompte, 1987; Mason, 1996; Marcus, 1998; Schwandt, 1997). En otras palabras, la **reflexión** es una autoobservación rigurosa llevada a cabo por el investigador a lo largo de todo el proceso de investigación. Se procesan o reconstruyen todos los datos a través de la mente del investigador tal y como aparecen escritos en el informe. La reflexión se considera un procedimiento muy importante para establecer la validez, especialmente entre los investigadores críticos que temen que se considere su trabajo empírico como un discurso ideológico o temerosos de la duplicación de preferencias sociales, raciales, étnicas y de género en sus estudios. En algunos estudios críticos la reflexión podría requerir estrategias adicionales para el rigor. Los investigadores cualitativos, en vez de negar la subjetividad humana, tienen en cuenta la subjetividad a través de estrategias metodológicas.

La subjetividad interpersonal

La investigación interactiva depende, en gran parte, de las habilidades interpersonales del que realiza la averiguación. Estas habilidades implican el establecimiento de buenas relaciones, ausencia de juicios, respeto a las normas de la situación y muestras de sensibilidad ante las cuestiones éticas. El investigador se relaciona con el participante como una persona, no como esponja inerte que «chupa» a su alrededor. Las emociones interpersonales en el campo de trabajo son esenciales en las actividades de recopilación de datos a causa de la interacción directa. Los sentimientos resultan de mucha utilidad durante el proceso de la investigación (Kleinman and Copp, 1993).

El progreso del estudio depende, a menudo, en primer lugar de la relación que el investigador establece con los participantes. Las habilidades esenciales son las que facilitan la conversación con los demás: ser un oyente activo, paciente y con capacidad de escucha y demostrar una comprensión empática y un respeto profundo a los puntos de vista de los participantes. El autocontrol del investigador está íntimamente ligado a la ética personal de cada uno. Además, el proceso interactivo es relativamente personal; ningún investigador observa, entrevista o se relaciona con los demás exactamente como lo haría otro. Estas cuestiones se tratan en primer lugar en el estudio para mejorar la reflexión.

Los datos obtenidos por los informadores son válidos aunque representen una opinión particular o hayan estado influenciados por la presencia del investigador. Estos datos son problemáticos sólo si se pide que los datos sean representativos fuera del contexto. Las posibles preferencias del investigador pueden minimizarse si los investigadores dedican el tiempo suficiente en el campo empleando estrategias de recopilación de datos múltiples para obtener datos a partir de enfoques diferentes. El espacio prolongado de tiempo para la recopilación de datos permite al investigador corroborar los datos e identificar la información no representada. La reacción del participante, la corroboración independiente y la confirmación—haciéndolo a lo largo de todas las etapas del proceso de investigación—son probablemente las técnicas más efectivas para identificar las preferencias del investigador. Es necesario aportar detalles suficientes sobre el diseño e incluir las estrategias seleccionadas para mejorar la subjetividad *disciplinada*.

Estrategias para mejorar la reflexión

Los investigadores cualitativos combinan cualquiera de las siete estrategias posibles para controlar y evaluar el impacto de su subjetividad y su punto de vista (ver tabla 10.4). Las estrategias más importantes mantienen un registro de campo y un diario y documentación de campo con capacidad de ser escuchada. Otras estrategias se añaden por resultar necesarias o requeridas para la validez de los datos.

Colega curioso. Un colega curioso es un compañero de trabajo desinteresado que se implica en las discusiones sobre los análisis preliminares del investigador y las posteriores estrategias metodológicas en un diseño emergente. Estas discusiones resaltan el conocimiento implícito que el investigador ha adquirido. El comentarista plantea, también, preguntas para ayudar al investigador a entender su propia postura y los valores y su función en la investigación. Además, las conversaciones podrían reducir la tensión que acompaña, normalmente, al trabajo de campo.

Registro de campo. Un registro de campo es una grabación cronológica, según fecha y horario, empleada en el campo y que afecta a los escenarios y a los informadores. El registro de campo contiene, también, para cada entrada los escenarios y las personas implicadas. El registro documenta el continuo trabajo

TABLA 10.4: ESTRATEGIAS PARA MEJORAR LA REFLEXIBILIDAD: ESTRATEGIAS PARA CONTROLAR Y EVALUAR LA SUBJETIVIDAD Y LA PERSPECTIVA DEL INVESTIGADOR

Estrategia	Descripción
Colega curioso	Selecciona un compañero que facilita el análisis lógico de los datos y la interpretación; realizada, a menudo, cuando el asunto posee mucha carga emocional o las experiencias del investigador son contrarias a los valores durante la recopilación de datos.
Registro de campo	Mantiene un registro de fechas, tiempo, escenarios, personas, y actividades para obtener el acceso a los informadores y para cada colección de datos recopilada.
Diario (inmediato) de campo	Registro de las decisiones tomadas durante el diseño emergente y la argumentación; incluye consideraciones sobre la validez de los datos.
Consideraciones éticas registradas	Registro de los conflictos éticos, las decisiones y las actividades en el diario de campo, y de las autorreflexiones.
Auditoría	Registro de las técnicas de organización de datos, los códigos, las categorías como una «serie de decisiones».
Corroboraciones formales la información inicial	Se encarga de las actividades de confirmación formal como una encuesta, grupos elegidos o entrevistas.
Reflexión crítica	Autocrítica planteando preguntas difíciles; punto de vista.

de campo esencial para la validez del diseño. Con frecuencia, un registro es un apéndice del estudio o se condensa en forma de diagrama para ilustrar las múltiples fuentes y situaciones.

El diario de campo. Un diario de campo es un registro continuo de las decisiones tomadas durante el diseño inicial y la argumentación en aquel momento. De manera que, el investigador puede justificar, basándose en la información disponible, las modificaciones y las reformulaciones del problema de la investigación y las estrategias. Estos registros constituyen la base para la presentación de los cambios en una sección metodológica de un estudio terminado. El diario de campo, también, podría contener afirmaciones de la autenticidad de cada conjunto de datos y de los informadores. Algunos diarios de campos, llamados periódicos reflejo, contienen la evolución de las ideas del investigador, sus reacciones personales, la bibliografía con posibles relaciones y las preguntas para la investigación posterior.

Consideraciones éticas grabadas. Los investigadores cualitativos, a menudo, se enfrentan a conflictos éticos y toman decisiones para resolverlos y continuar el estudio. Se mantiene una grabación de las decisiones, de las personas implicadas, de las decisiones adoptadas, y del impacto en el estudio. Los investigadores eligen estrategias en el campo; algunas decisiones se basan, sobre todo, en consideraciones éticas. Mantener una grabación de consideraciones éticas ayuda a justificar las posteriores decisiones.

Auditoría. La auditoría es la conservación de los registros de las técnicas empleadas en el manejo de los datos y de las decisiones —normas que documentan la «cadena de evidencias» o la «serie de decisiones»—. En la investigación cualitativa, la integridad de la información se basa en los propios datos. Se mantiene un registro que relaciona la información con sus fuentes originales o con los datos. El registro incluye las técnicas de organización de datos y los códigos, las categorías y los temas utilizados para construir e interpretar el caso narrativo. El registro incluye esbozos y los preliminares diagramas integrantes. De este modo, la cadena de evidencia está disponible para que revisores externos la inspeccionen y la corroboren.

Los criterios para la auditoría pueden encontrarse con o sin un revisor externo, es decir, un «auditor». A menudo, no hay ninguno disponible que pueda servir como revisor externo porque este revisor debe conocer tanto la metodología como el asunto de la investigación. Una técnica alternativa es incluir una lista de documentos, códigos, categorías y de decisiones y normas en un apéndice para que el lector lo lea atentamente. Si se realiza una auditoría formal, el revisor externo comprueba la precisión de los documentos que vinculan la información con las fuentes originales y aportan una opinión del revisor.

Corroboración formal de la información inicial. Cuando la información depende de algunos informadores decisivos, los investigadores podrían realizar la confirmación formal de las actividades de recopilación de datos, como una encuesta, entre grupos seleccionados y entrevistas en profundidad con individuos con abundante información y que no fueron seleccionados en un primer momento de la investigación. Para advertir que se trata de una actividad de corroboración, lo primero que debe hacerse es analizar íntegramente los datos cualitativos. Las actividades de confirmación aseguran que los patrones acerca de los datos no están contaminados en exceso por la subjetividad del investigador. Las actividades de confirmación, si son posibles, se añaden a la credibilidad del estudio.

REFLEXIÓN CRÍTICA

La reflexión requiere que el investigador haga algo más que autocontrolarse y mantener un registro de las decisiones. La reflexión crítica es la confirmación constante, por parte del investigador, de sus actividades y funciones durante todo

el proceso de la investigación. Esta concienciación fomentada promueve la transformación social y personal. El propio acto de plantearse preguntas difíciles a uno mismo supone que el investigador no puede ser neutral, objetivo o independiente. Por ejemplo una pregunta difícil está relacionada con la « voz » –se permitió que emergieran todas las voces, especialmente las silenciadas socialmente porque sus puntos de vista, a menudo, son tenidos en cuenta por la situación.

Otra estrategia de reflejo es el **posicionamiento** que asume que sólo los textos en los que los investigadores expresan sus propias posturas y argumentos contextuales para razonar podrían considerarse «correcta» investigación (Lincoln, 1995; Lather, 1991; J. K. Smith, 1993). Los investigadores críticos, con frecuencia, escriben en la introducción su posicionamiento social, cultural, histórico, racial y sexual dentro del estudio. Por ejemplo, en la introducción *I Answer with MY life: Life Histories of Women Teachers Working for Social Change*, K. Casey (1993) describía su propia identidad como profesora que trabajaba por el cambio social. Esto incluía sus preferencias metodológicas, la historia de su vida y añadía sus puntos de vista sobre la presentación de los datos y el análisis. Sin embargo, las consideraciones de posicionamiento pueden ser sólo un gesto. En la investigación de la acción participativa, el investigador escribe en los datos sus propias acciones. Una función complicada del investigador y del participante requiere una inspección tanto de las funciones como de los datos resultantes. Esta no es una tarea fácil y no se le puede restar importancia.

EXTENSIÓN DE LA INFORMACIÓN CUALITATIVA

Los investigadores cualitativos elaboran auténticas descripciones analíticas de fenómenos que son informativos y útiles para la investigación de comunidades, lectores y participantes.

Cuestiones de autenticidad y utilidad

La mayoría de los estudios cualitativos emplean un diseño de estudio de caso en el que el caso no se trata como una muestra de probabilidad del universo más amplio. En otras palabras, el investigador no pretende generalizar los resultados. Sin embargo, lo que se busca es conseguir la **extensión de los descubrimientos**, que permite a los demás entender situaciones similares y aplicar estas informaciones a investigaciones futuras o a situaciones prácticas. El conocimiento no se produce mediante una réplica sino mediante la preponderancia de la evidencia encontrada en estudios de caso separados a través del tiempo o en diseños cuantitativos más estructurados.

La **autenticidad** es la fiel reconstrucción de las múltiples impresiones de los participantes. Para ser útil, el diseño de la investigación debe ser descrito de

manera adecuada. Algunos estudios, tales como el argumento de una teoría, citan los esbozos teóricos y las estrategias de investigación que otros investigadores entienden. Otros estudios son más contextuales y se escriben por las implicaciones prácticas. Un estudio que es una investigación sistemática pero idiosincrática posee una relevancia limitada para la investigación futura.

Componentes del diseño para generar la extensión de los descubrimientos

Diez componentes del diseño afectan a las extensiones lógicas (no estadísticas): la función de la investigación, la selección de los informadores, el contexto social, las estrategias de recopilación de datos, las estrategias de análisis de datos, el relato auténtico, la tipicidad, las premisas analíticas, las explicaciones alternativas, y los criterios asociados con un propósito particular de investigación (ver tabla 10.5).

Función de la investigación. La importancia de la relación social del investigador con los participantes requiere que los estudios identifiquen la función del investigador y su situación dentro del grupo. La función del investigador preferente es la de una persona que es desconocida en el lugar o para los participantes, en otras palabras, un «extraño». Los investigadores cualitativos citan también con frecuencia experiencias personales o profesionales que les permiten relacionarse con los participantes, es decir, identifican mejor los procesos observados y los significados sutiles de los participantes que los que carecen de tales experiencias. Sin embargo, la investigación de acción participativa requiere la función del participante y del investigador.

Selección del informador. La selección del informador se describe con la justificación de la selección y el proceso de decisiones. La investigación posterior requiere que un investigador contacte con individuos parecidos a los que fueron informadores en el estudio anterior.

Los investigadores cualitativos ven la pérdida de sujetos como un acontecimiento normal. Valoran a cada «persona»; los informadores y los participantes no son intercambiables. Si un informador no es cooperativo o no está disponible, los investigadores continúan buscando otros informadores esenciales o utilizan la evidencia de la documentación hasta que aparece un patrón.

Aumentar el número de escenarios en un diseño de estudio de los casos no facilita la ampliación de la información; los escenarios no son muestras probables.

Contexto social. Los contextos sociales influyen sobre el contenido de los datos y son descritos materialmente, socialmente, interpersonalmente y funcionalmente. Las descripciones físicas de las personas, el momento y el lugar de los acontecimientos o de las entrevistas prestan ayuda en el análisis de los datos. Las relaciones interpersonales entre los miembros del grupo y las relaciones sociales

TABLA 10.5: COMPONENTES DEL DISEÑO PARA GENERAR EL USO DEL ESTUDIO Y LA EXTENSIÓN DE LA INFORMACIÓN

Estrategia	Descripción adecuada para el estudio
Función de la investigación	La relación social del investigador con los participantes.
Selección del informador	Los criterios, la justificación y el proceso de decisiones utilizado en el muestreo intencionado.
Contexto social	Los escenarios sociales, físicos, interpersonales y funcionales de la recopilación de datos.
Estrategias de recopilación	La variedad de métodos empleada, incluyendo la observación del participante, la entrevista, los documentos y otros.
Estrategia de análisis de datos	El proceso de análisis de datos descriptivos.
Relato auténtico	La extensa descripción presentada como un relato analítico.
Tipicidad	Las características distintivas de grupos y/o de lugares presentados.
Premisas analíticas	El esbozo teórico o normativo inicial que da forma al estudio.
Explicaciones alternativas	El esquema retrospectivo de todas las explicaciones contrarias o verosímiles para las interpretaciones.
Otros criterios para el enfoque de la investigación (tras haber completado el estudio)	
Etnografía	La explicación global de la complejidad de la vida del grupo.
Fenomenología	Entender la esencia de la experiencia vivida; genera más preguntas de investigación.
Estudio de caso	Entender la práctica; facilita la toma de decisiones.
Argumento de la teoría	Los conceptos o las propuestas relacionadas con la ciencia social; genera la investigación de verificación con diseños más estructurados.
Tradiciones críticas	Informa o anima a los participantes sobre su situación y oportunidades; genera una investigación más amplia; estímulo de acción.

entre los grupos podrían explicar las actuaciones individuales y sus significados. El propósito de los encuentros del grupo, las conferencias de los padres o los

encuentros oficiales requieren otros investigadores para encontrar contextos sociales similares para un estudio más amplio.

Estrategias de recopilación de datos. Usar el estudio para la futura investigación es imposible sin precisar las descripciones de las técnicas de recopilación de datos: las variedades de métodos de observación y de entrevistas y las técnicas de registro de datos con las circunstancias y los procesos para asegurar registros precisos. Se observa cómo se emplean las diferentes estrategias. Por ejemplo, si el interés residía en la observación de un proceso de grupo, entonces las entrevistas individuales y los documentos servirían como estrategias de corroboración de la recopilación de datos.

Estrategias de análisis de datos. Simplemente afirmar que el análisis formal de datos se hizo minuciosamente es insuficiente para facilitar la extensión de la información. El investigador debe aportar información retrospectiva de cómo se sintetizaron los datos y de cómo se identificaron las estrategias generales del análisis de los datos y la interpretación. Con frecuencia las categorías utilizadas en el análisis de los datos y sus normas sobre decisiones aparecen enumeradas en un apéndice.

Relato auténtico. La mayoría de los estudios cualitativos contiene una extensa descripción en un relato analítico repleto de largas citas que reflejan el lenguaje de los participantes.

Un buen relato es el que podría leerse y vivirse indirectamente por otros. Los posibles relatos facilitan significados que dan sentido y definen el pensamiento de los lectores y las prácticas. Un relato resulta auténtico cuando los lectores conectan con la historia reconociendo peculiaridades, imaginando las escenas y reconstruyéndolas a partir de recuerdos asociados con acontecimientos similares (Connelly y Clandinin, 1990). Es lo particular y no lo general lo que influye sobre las personas y sus emociones. Las historias se encuentran entre lo general y lo particular a partir de los requisitos básicos de la investigación y los requisitos de vida personales prácticos y concretos.

Tipicidad. La extensión de la tipicidad del fenómeno (Wolcott, 1973), es decir, presenta el grado en el que podría compararse o contrastarse a lo largo de dimensiones relevantes con otros fenómenos. La «obsesión virtual» de los investigadores cualitativos por la descripción de las diferentes características de los grupos estudiados demuestra la valoración de la importancia de este tipo de información para la ampliación de objetivos.

Tanto los atributos cualitativos como los cuantitativos de los grupos y de los escenarios son esenciales. Incluyen el ambiente socioeconómico, el logro educativo, el grupo de edad, la composición étnica o racial, el periodo temporal y los rasgos contextuales de localización. Las experiencias únicas históricas de grupos y culturas podría limitar la extensión, pero pocas experiencias de grupo son etnocéntricas. El extracto 10.4 describe la tipicidad de una escuela judía elemental

de dimensiones relevantes. Una vez que se establece la tipicidad, resulta evidente una base para la extensión de los conceptos entendidos, y la información aporta descubrimientos a lo largo del tiempo o de las situaciones.

EXTRACTO 10.4 **TIPICIDAD**

La escuela particular elegida para el estudio era una escuela elemental bajo el control de una sinagoga judía conservadora. Se deseaba este tipo de escuela judía porque es estadísticamente típica entre un amplio porcentaje de escuelas judías de Estados Unidos (Lang 1968). Como en el caso de esta escuela, el porcentaje mayor de alumnos que acudía a escuelas judías a nivel nacional (44.4 por ciento) asistía a la escuela entre 2 y 5 tardes por semana (Rockowitz y Lang 1976). Además, el mayor número de escuelas parece acoger entre 100 y 299 estudiantes, como sucede en esta escuela con sus aproximadamente 250 estudiantes (Rockowitz y Lang 1976). Finalmente, el mayor número de congregaciones conservadoras que regentaba estas escuelas poseía un número de miembros que oscilaba entre 100 y 249 familias (Friedman 1979), como era el caso de este estudio en el que la congregación tenía aproximadamente 200 familias registradas como miembros. La escuela funcionaba dos tardes (1'5 horas cada clase) y un domingo por la mañana (de 2'5 horas cada clase) cada semana, aunque una de las clases por la tarde era opcional. En general, el currículum coincide con el currículum estándar de la escuela por la tarde.

Fuente: David Schoen, (1982), «Explaining Jews Student Failure», *Anthropology & Education Quarterly*, 13(4). Reimpreso con permiso de *American Anthropological Association*.

Premisas analíticas. El componente del diseño primario que genera una investigación más amplia es un esbozo conceptual explícito que compone el estudio y con el que la información puede integrarse o contrastarse a partir de una investigación anterior. La selección de un esbozo conceptual para un estudio requiere que otros investigadores comiencen, también, a partir de premisas analíticas similares. Ver el extracto 10.5 para las consideraciones explícitas o las premisas analíticas.

Como la principal consecuencia de la investigación cualitativa es la generación y el refinamiento de los conceptos, los investigadores cualitativos deben contrastar su información con la investigación anterior. Cuando aparecen discrepancias, los investigadores cualitativos presentan los atributos del grupo, el periodo temporal y las situaciones. Esto alerta a otros investigadores cuando usan esta información.

Sin embargo, muchas investigaciones cualitativas poseen un objetivo primordialmente descriptivo y explicativo y no se limitan a los estudios teóricos. La decisión de centrarse en la descripción narrativa y analítica está, en parte,

relacionada con la investigación anterior y con la finalidad de la investigación. De este modo, el párrafo final de un estudio de un profesor bilingüe (extracto 10.6) afirma que esperaba que los lectores estuvieran más informados sobre la educación bilingüe basada en la realidad de sus experiencias.

Explicaciones alternativas. Durante el análisis de los datos, los investigadores cualitativos buscan evidencias negativas o datos discrepantes para intercambiar o modificar los patrones iniciales.

EXTRACTO 10.5 **PREMISAS ANALÍTICAS**

Por lo tanto, no es extraño que las chicas cuyas familias no eran como las familias de las series de televisión, consideraran que las familias de las series eran menos reales. Según el enfoque cultural de McRobbie (1991), las chicas estaban interpretando y enjuiciando medios de comunicación en base a sus propias experiencias personales. Greenberg y Reeves (1976) encuentran también que las experiencias personales de los niños afectan a sus percepciones de la realidad en la televisión. Los niños, en general, no tienen múltiples puntos de referencia con respecto a la vida familiar, de manera que las chicas de este estudio no veían ninguna familia diferente de las que consideraban surrealistas. Estos datos también corroboran de manera empírica la predicción de Greenberg y Reeves sobre que los contenidos de la televisión se perciben como más parecidos a la vida real si las actitudes y comportamientos del niño, o en este caso, la dinámica de la familia, concuerdan con el contenido televisivo.

Fuente: Fingerson, L. (1999), «Active viwing. Girls' interpretations of family television prgrams» (p. 398-399). *Journal of Contemporary Ethnography*, 28(4), 389-4128.

EXTRACTO 10.6 **LA EXTENSIÓN DE LA INVESTIGACIÓN DESCRIPTIVA Y ANALÍTICA**

Estas estrategias la conducen (la señorita S.) de forma implícita, aplicando unos puntos de vista acerca de los procesos educativos, tales como que ellos pertenecen a una minoría de estudiantes. Una investigación desde este enfoque, que se desarrolla aparte de las experiencias comparables de los profesores en este país (Estados Unidos), es una terrible necesidad en el campo de la reflexión educativa. Este tipo de investigación podría llegar a dirigir un tipo de entendimiento de educación bilingüe que es más sofisticado, y tiene más en cuenta los matices y las diversidades de las comunidades minoritarias –sobre todo, más informadas por las *realidades* de su experiencia–, que lo que este caso presenta (p.189).

Fuente: Montero-Sieburth, M. y Pérez, M. (1987), «Echar Pa'lante., Moving Onward:The Dilemmasand Strategiesof a Bilingual Teacher, *Anthopology & Education Quarterly*, 18 (3), 189. Reimpreso con permiso de *American Anbthropological Association*.

Los etnógrafos y los entrevistadores buscan de manera activa a los informadores y los escenarios sociales que parecen variar u oponerse a los datos anteriores. Un patrón principal se convierte en una explicación, sólo, cuando los modelos alternativos no ofrecen explicaciones razonables al problema de la investigación. Las explicaciones alternativas y principales se discuten en el estudio porque ambas deberían dar lugar a una investigación más amplia.

Otros criterios para el enfoque de la investigación (tras haber completado el estudio). Además de aportar una descripción adecuada de los componentes generales mencionados, los enfoques específicos de la investigación podrían hacer hincapié en criterios adicionales. Por ejemplo, la fenomenología, las teorías argumentadas y los estudios críticos podrían tener efectos algo diferentes sobre las comunidades de investigación, los lectores y los participantes. La etnografía ofrece una comprensión global de la complejidad de la vida de grupo, que conduce a estudios más amplios de casos. Un estudio fenomenológico aumenta la comprensión de las experiencias vividas por los lectores y por otros. Un estudio de caso consigue un mejor entendimiento de la práctica o de la cuestión y facilita la toma de decisiones con información. Sin embargo, a menudo, el estudio de la teoría argumentada conduce a diseños más estructurados para examinar el concepto o verificar una afirmación. Algunas formas de tradiciones críticas no sólo informan a través de un proceso histórico de revisión, sino que refuerzan y estimulan la acción. Por consiguiente, el investigador se reúne, con frecuencia, con los informadores principales o con todos los participantes, revisa la información y establece un diálogo. Los investigadores podrían, también, aportar fuentes adicionales de información para adoptar un refuerzo personal y de grupo.

Cuando los investigadores cualitativos han tratado, de manera apropiada, las cuestiones de validez del diseño, la subjetividad disciplinada y la extensión de los descubrimientos señaladas anteriormente, otros investigadores cualitativos consideran su trabajo, a menudo, creíble. Muchas cuestiones de diseño se tratan mediante la organización y el control de estudios basados en los criterios apropiados para la credibilidad de la investigación.

ÉTICA DE LA INVESTIGACIÓN: FUNCIONES Y RECIPROCIDAD

Los principios éticos en la investigación cualitativa son parecidos a los de la investigación cuantitativa. Las directrices éticas incluyen, pero no están limitadas por ello, informe de consentimiento, decepción, confidencialidad, anonimato, daño a sujetos, privacidad y otros (ver el capítulo 3). La *American Anthropological Association*, la *American Psychological Association*, la *American Educational Research Association*, y la *American Sociological Association* poseen directrices éticas profesionales y patrones de actuación. Si embargo, los trabajadores de campo deben adoptar estos principios en situaciones complejas.

Dilemas éticos en el campo de trabajo

Los investigadores cualitativos tienen que ser sensibles a los principios éticos a causa de su tema de investigación, la recopilación directa de datos interactivos, un diseño emergente y la reciprocidad con los participantes. Los criterios para un diseño de investigación incluyen, no sólo, la selección de informadores con abundante información y las estrategias eficaces de investigación, sino también, la adhesión a la ética de la investigación.

Por ejemplo, algunos investigadores cualitativos se han ganado la confianza de personas potencialmente implicadas en actividades ilegales para recopilar datos. Otros investigadores han investigado asuntos sensibles políticamente o que generan controversia. Cuando un investigador estudia a aquéllos cuyos actos se consideran criminales o abusivos, surge un profundo dilema ético: «cuando alguien decide intentar entrar en su mundo y estudiarlo, el trabajador de campo llega a una crisis moral y existencial» (Soloway y Walter, 1977, p.161). Los investigadores se han enfrentado a dilemas éticos al estudiar la adicción a las drogas, la brigada antidroga y los cuidados en las instituciones para personas con discapacidades psíquicas. Como los investigadores cualitativos quedan implicados en las situaciones y la vida diaria de los participantes, estos investigadores se encuentran, a menudo, con situaciones moralmente problemáticas.

Los diseños iniciales requieren que se consideren los principios éticos a lo largo de todas las fases de la organización y la recopilación de los datos. Además, algunas decisiones tomadas en el momento tienen implicaciones éticas (Taylor, 1987). Aquí aparecen algunas preguntas típicas: «¿observo este abuso o lo ignoro?» «¿Grabo esta confesión?» y, «si la grabo, ¿la incluyo en un informe público?» «Si observo un abuso o negligencia, ¿informo a la policía?» «¿Estoy realmente observando un abuso o negligencia o proyectando mis valores en la situación?» «Si prometí confidencialidad cuando me introduje en el campo como investigador, estoy rompiendo mi “juramento” si interfiero?» «Si rompo mi “juramento”, se me pedirá que abandone el campo; si lo abandono, quedará toda la comunidad pública, profesional y de investigación desinformada?». Estas preguntas sugieren que resulta difícil separar la ética de la investigación de la ética profesional y de la moral personal.

La mayoría de los investigadores cualitativos idean funciones que provocan la cooperación, la apertura y la aceptación. A veces, esto significa que los investigadores asumen una función de ayuda, dirigida en un cierto sentido, o aceptan ser manipulados.

Cuando la gente adapta sus prioridades y su rutina para ayudar a un investigador o, incluso, tolerar su presencia, están dando algo de sí mismos. Un investigador está agradecido con estas personas. Algunos estudiantes (Marshall y Rossman, 1999) sugieren que los investigadores ingenien maneras, dentro de las

limitaciones de la investigación y de la ética personal, para que sea recíproco. La reciprocidad puede deberse al tiempo, a los conocimientos, a la atención o a los servicios especializados. Algunos investigadores, al final del informe, se convierten en abogados de un grupo particular de la amplia comunidad, incluyendo los grupos que elaboran las normas.

Ética de la investigación en el campo de trabajo

La mayoría de las situaciones éticas requiere que los investigadores determinen las prioridades de situación que, a menudo, implican discrepancias con los participantes. A continuación, aparecerán conflictos éticos que pueden plantearse durante la realización del trabajo de campo y la manera en que han sido resueltos por los investigadores.

Informe de consentimiento como diálogo. Al obtener el permiso para entrar en el campo, la mayoría de los investigadores aseguran la confidencialidad y el anonimato y describen el uso pretendido de los datos. A causa de las estrategias de investigación empleadas, raramente, surge la cuestión del daño psicológico o físico.

Muchos investigadores (L. M. Smith, 1990) consideran el informe de consentimiento como un diálogo: cada nuevo participante en el estudio es informado del propósito y se le asegura la confidencialidad y el anonimato. Sin embargo, se presentan situaciones en las que resulta imposible un diálogo, tales como un trauma repentino inesperado que lleva a las personas a la escena durante la observación, o la observación del comportamiento público, como una muchedumbre durante un partido de fútbol. Se informa a los participantes de manera que se anime así su libre elección de participación. Se explica sobre el tiempo requerido para la participación y sobre la función de la investigación que no interfiere ni juzga. Con frecuencia, los informadores seleccionan los escenarios y el momento de la entrevista. Dado que los investigadores intentan establecer relaciones auténticas, necesitan planear cómo establecer el diálogo. La mayoría de los participantes detectan y rechazan la falta de sinceridad y la manipulación.

Confidencialidad y anonimato. En general, existe un fuerte sentimiento entre los trabajadores de campo de que las situaciones y los participantes no deberían resultar identificables.

Los investigadores utilizan localizaciones imaginarias y disfrazan los rasgos de las situaciones para hacerlas parecer similares a otros escenarios posibles. Los investigadores, a menudo, codifican los nombres de las personas y de los lugares. La revisión del participante en el estudio de caso de cada individuo sirve para asegurar la confidencialidad y para proteger la privacidad. Además, la policía puede revisar un informe antes de que éste se considere como terminado. Los investigadores tienen una doble responsabilidad: la protección de las confidencias

de los participantes de otras personas en la situación cuya información privada podría permitirles su identificación, y la protección de los informadores del público general de lectores.

Sin embargo, la ley no protege a los investigadores si el gobierno les obliga a hacer públicos asuntos confidenciales. El informe, las notas de campo y el investigador pueden ser castigados por la justicia. Por ejemplo, un investigador ha completado prácticamente un trabajo de campo, cuando un abogado escolar afirmó que era un testigo conocedor de un caso de disgregación escolar. El investigador inició el «principio ético de diálogo» (L. M. Smith, 1990, p. 271) al presentar el dilema a varios encargados para que resolvieran el problema. Finalmente, un alto funcionario anunció al abogado escolar que no llamaría al investigador como testigo porque, si lo hacía, violaba los compromisos de confidencialidad que hizo a la plantilla, a los profesores y a los directores del colegio. Spradley (1979) hace hincapié en la protección de las notas de campo e, incluso, sugiere que se considere un plan alternativo si la protección de los informadores no es posible.

Decepción, privacidad y refuerzo. La mayoría de los investigadores cualitativos ven la decepción cómo una violación del informe de consentimiento y de la privacidad. Sin embargo, algunos etnógrafos reputados se han presentado como vagabundos e, incluso, como reclutadores del ejército para recopilar datos sin informar a los participantes (Punch, 1986). Estos investigadores exigen que no se proporcione ningún daño a los informadores por la investigación. Incluso, las personas informadas que cooperan podrían sentirse traicionadas cuando leyese la información impresa.

Otros investigadores sugieren que los trabajadores de campo negocien con los participantes para que entiendan el poder que tienen en el proceso de la investigación. Este poder y las decisiones que conlleva podrían convertirse en un intercambio entre la privacidad perdida por la participación en un estudio (Lincoln 1990). Además, si los problemas éticos surgen en el trabajo de campo, los problemas recíprocos resueltos por el investigador y los participantes pueden, con frecuencia, conducir a alternativas. Un enfoque como éste insiste en considerar que, tanto las actividades de los participantes, como la investigación son importantes.

Daño, cuidado y justicia. Aunque el daño físico a los informadores se produce raramente en la investigación cualitativa, algunas personas pueden experimentar la humillación personal y la pérdida de la confianza interpersonal. Algunos investigadores (Cassell, 1982) formulan el principio de que las personas sean tratadas como fines en sí mismos en lugar de cómo medios para llegar a un final, justificando el posible daño a un individuo porque esto podría ayudar a otros. Otros investigadores consideran que un sentido de preocupación y de justicia tiene que incluirse en los pensamientos y en las acciones del investigador. La ética profesional y de la investigación para muchos investigadores está íntimamente relacionada con la moral personal.

Muchos investigadores argumentan «un relativismo determinado» o «racionalización» en una situación concreta. Aunque existen las directrices éticas, la dificultad está en su aplicación. Los investigadores necesitan identificar los potenciales conflictos éticos y resolverlos. A menudo, las discusiones abiertas y la negociación fomentan la «imparcialidad» de las personas y de la búsqueda de la investigación. El conocimiento puede servir tanto para animar a los participantes en sus situaciones como para la comunidad profesional más amplia.

NORMAS DE ADECUACIÓN

Los diseños cualitativos se juzgan mediante varios criterios. A continuación aparecerán las preguntas típicas que los investigadores podrían plantear a sus diseños o que los revisores podrían utilizar para criticar un diseño cualitativo.

1. ¿Debe ser estudiado el fenómeno claramente articulado y delimitado?
2. ¿Se presentan las preguntas de la investigación, el propósito, el modo de averiguación y el diseño?
3. ¿Qué técnicas de muestreo intencionado se emplearán para identificar los casos de información abundante? ¿Parece probable que la estrategia de determinación de modelos obtenga grupos o casos con abundante información? (normalmente es necesaria una información preliminar antes de que la estrategia de la determinación de modelos pueda elegirse).
4. ¿Se presenta el tamaño mínimo deseado de la muestra? ¿Parece el tamaño de la muestra lógico para aportar datos abundantes sobre el fenómeno en un espacio de tiempo razonable?
5. ¿Se presenta el diseño con detalles suficientes para mejorar la validez; es decir, se especifican las estrategias esenciales como el trabajo de campo prolongado, la recopilación de repeticiones con datos descriptivos y la búsqueda de casos negativos?
6. ¿Qué estrategias de recopilación de datos múltiples se espera que aumenten el acuerdo en la descripción del fenómeno y sus significados entre el investigador y los participantes? ¿Posee el investigador conocimiento y experiencia con las estrategias propuestas o tiene un estudio preliminar?
7. ¿Sugiere el diseño la naturaleza inicial del estudio?
8. ¿Qué estrategias proyecta emplear el investigador para mejorar la reflexión?
9. ¿Qué componentes del diseño se incluyen para fomentar la utilidad y la extensión lógica de los descubrimientos? ¿Existen otros que podrían emplearse, y si es así, cuáles son?
10. ¿Especifica el investigador cómo se tratarán en el campo el informe de consentimiento, la confidencialidad, el anonimato y otros principios éticos?

Los diseños cualitativos de la investigación son, con frecuencia, difíciles de juzgar a causa de la reflexión y de la naturaleza inicial del diseño. Los diseños, si son realmente iniciales y para investigar, se modificarán a medida que avanza el estudio. Muchos de los patrones están relacionados con las estrategias de la recopilación de datos. (ver *Normas de adecuación para estrategias cualitativas* en el capítulo 11).

RESUMEN

Las siguientes frases resumen los aspectos principales del diseño de investigación cualitativa.

1. Los investigadores cualitativos estudian las situaciones de los participantes: sentimientos, pensamientos, opiniones, ideales y, también, las actuaciones en situaciones normales.
2. Los investigadores cualitativos utilizan estrategias interactivas para recopilar datos destinados a los estudios de exploración, explicación o emancipatorios.
3. Los investigadores cualitativos emplean un diseño emergente.
4. Un diseño de un estudio de caso se centra en un fenómeno para entenderlo en profundidad, independientemente de las personas o escenarios del estudio.
5. Los estudios de caso son significativos para la práctica de la teoría, las normas y el desarrollo de la acción social.
6. El muestreo intencionado selecciona muestras reducidas de casos con abundante información para estudiar profundamente sin el deseo de generalizar a todos los casos.
7. Los tipos de muestreo intencionado incluyen la selección del escenario, el muestreo comprensible, el muestreo de variación máxima, el muestreo en cascada y el muestreo mediante tipo de caso.
8. El tamaño de la muestra depende del propósito del estudio, de las estrategias de recopilación de datos y de la disponibilidad de los casos con abundante información.
9. La recopilación de datos y el análisis son interactivos y ocurren en ciclos que se superponen.
10. El empleo de estrategias de investigación en vez de procedimientos permite la flexibilidad en el estudio y la corroboración de cada nueva idea tal y como ocurre en la recopilación de datos.
11. Las fases de la investigación cualitativa son organización, recopilación de datos iniciales, recopilación de datos básicos y diagramas y análisis formal de datos.

12. La validez de los diseños cualitativos es el grado en el que las interpretaciones y los conceptos tienen significados recíprocos entre los participantes y el investigador.
13. Los investigadores cualitativos mejoran la validez haciendo explícitos todos los aspectos de sus diseños.
14. Las estrategias de recopilación de datos para aumentar la validez son una combinación de las estrategias siguientes: indicadores, investigadores múltiples, datos registrados mecánicamente, investigador participante, comprobación de los miembros, revisión del participante y presentación de casos negativos.
15. Los investigadores cualitativos emplean la subjetividad interpersonal para recopilar datos y estrategias de reflexión para mejorar el estudio.
16. Los estudios cualitativos pretenden una ampliación de la información en lugar de la generalización de los resultados. Normalmente, la generalizabilidad no es el objetivo del estudio.
17. Los componentes del diseño que mejoran la extensión de los descubrimientos son la especificación de la función del investigador, la selección del informador, el contexto social, la recopilación de datos y las estrategias de análisis, el relato auténtico, la tipicidad, las premisas analíticas y las explicaciones alternativas.
18. Los investigadores de campo, con frecuencia, emplean el diálogo y la reciprocidad mientras actúan según principios éticos y legales con los participantes.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. El propósito de un diseño de estudio de caso es...
 - a. generalizar a casos similares.
 - b. conseguir una respuesta creíble a un problema.
 - c. obtener la comprensión profunda de «un» fenómeno.
 - d. seleccionar un lugar para la observación.
2. Los estudios de caso no...
 - a. desarrollan una teoría.
 - b. comparan grupos estadísticamente en diferentes condiciones experimentales.
 - c. describen y analizan una práctica o un proceso.
 - d. identifican las normas.
3. La lógica del muestreo intencionado es que...

- a. un modelo representativo de forma aleatoria puede generalizarse a una población.
 - b. algunos casos, con abundante información, estudiados en profundidad aportan mucha información sobre un tema.
 - c. se incluyen todos los casos, incluso, si no son relevantes con respecto al problema.
 - d. los casos disponibles se utilizan porque el investigador tiene acceso a ellos.
4. ¿Cuál de las respuestas siguientes describe correctamente las fases de la investigación cualitativa?
- a. identificar el problema de la investigación, organización, recopilación de datos, análisis de datos, finalización.
 - b. organización, recopilación de datos iniciales, recopilación de datos básicos, recopilación de datos finales, finalización.
 - c. identificación de las hipótesis de la investigación, recopilación de datos, análisis de datos, conclusiones.
 - d. organización, identificación del problema, recopilación de datos, finalización.
5. Las estrategias siguientes mejoran la validez del diseño «excepto»...
- a. utilizar indicadores de baja inferencia.
 - b. obtener repeticiones.
 - c. conducir la revisión de los participantes.
 - d. ignorar los casos negativos.
6. La validez del diseño cualitativo es...
- a. aumentada en un periodo muy breve de recopilación de datos.
 - b. mejorada mediante el uso de un lenguaje social científico abstracto en lugar del lenguaje de los participantes.
 - c. mejorada con la comprobación de los miembros
 - d. aumentada mediante la interferencia del investigador en los acontecimientos habituales en el escenario.
7. La validez del diseño en la investigación cualitativa se refiere a...
- a. la consistencia de las afirmaciones.
 - b. con qué facilidad las descripciones y las interpretaciones poseen significados recíprocos entre el investigador y los participantes.
 - c. el empleo del modelo aleatorio.
 - d. el empleo de una variable experimental.
8. La reflexión y la subjetividad disciplinadas reconocen que...
- a. las emociones en el trabajo de campos son esenciales en las interacciones directas pero deben controlarse y evaluarse.
 - b. la mayoría de los investigadores cualitativos son imparciales y objetivos.

- c. no tiene importancia un registro de las decisiones de campo.
 - d. no es necesario establecer la audibilidad.
9. La extensión de la información en la investigación cualitativa...
- a. pretende una generalización de los resultados a otras poblaciones y situaciones.
 - b. depende de réplicas exactas del estudio de caso en otras situaciones.
 - c. es mejorada mediante la tipicidad, el relato auténtico y la especificación de todos los componentes del diseño.
 - d. es imposible de alcanzar con las estrategias cualitativas.
10. La ética de la investigación cualitativa incluye...
- a. el uso de los códigos y nombres falsos para mantener el anonimato.
 - b. el informe de consentimiento como un diálogo mediante el que las personas tienen poder sobre el proceso de la investigación.
 - c. la negociación con los participantes y la resolución mutua de conflictos éticos que tienen lugar en el campo.
 - d. todo lo anterior.

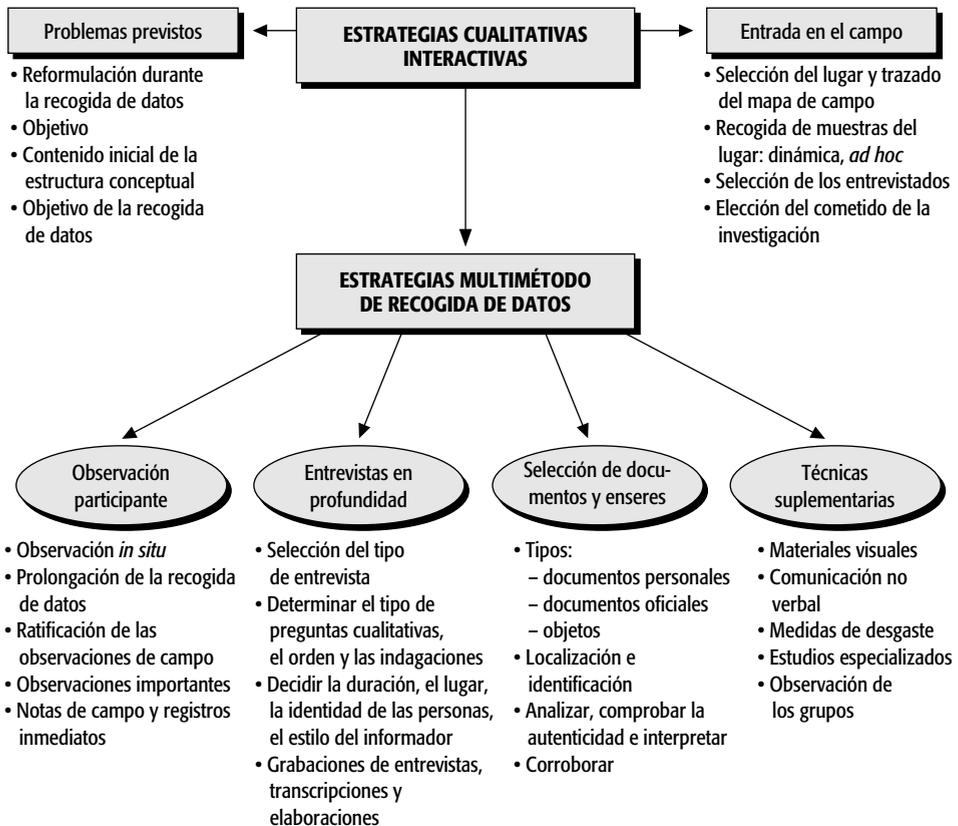
Problemas de aplicación

1. La directora de un programa de preescolar de una ciudad quiere obtener las opiniones de los padres acerca del programa. Está especialmente interesada en la innovación de este año en una clase: un programa educativo para los padres. Los materiales son enviados a las casas por un padre o un cuidador dos veces por semana para trabajar con el niño, y se graban los contactos entre padres y profesores. Hay doce niños en la clase de preescolar. Cuatro niños viven con uno sólo de los padres, seis viven con los dos y dos niños viven con un padre y sus abuelos.
 - a. ¿Qué tipo de determinación de modelos sería la apropiada y por qué? (probabilidad y propósito)
 - b. ¿Cómo debería hacerse la determinación de modelos?
 - c. ¿Cuáles serían los métodos cualitativos apropiados?
2. Un investigador está interesado en cómo los directores toman decisiones sobre los alumnos de escuela elemental que no alcanzan el nivel de su curso. ¿Cómo diseñarías tú este estudio?
3. Un investigador quiere entender el concepto de dirección de la escuela basada en la localización. Se ha establecido en un distrito escolar en el que ha pasado un año planeando y escribiendo las directrices para la dirección basada en la localización de las seis escuelas seleccionadas. El investigador está interesado, en primer lugar, en cómo funciona un equipo de dirección que se basa en la localización y en evaluar si esto afecta a la función del director y a las relaciones de la escuela con la dirección cen-

tral del distrito. Un equipo de dirección basado en la localización consiste en seis miembros de la comunidad y tres profesores con el director como miembro *ad hoc*. Un asistente de la oficina central atiende, con frecuencia, los encuentros mensuales tras haber dirigido seis sesiones orientativas. ¿Cómo diseñarías el estudio?

- a. ¿Qué tipo de determinación de modelos sería apropiada y por qué?
 - b. ¿Cómo debería hacerse la determinación de modelos?
 - c. ¿Qué métodos cualitativos serían apropiados?
4. Para el problema 1, ¿qué estrategias de investigación aumentarían la validez del diseño, minimizarían la parcialidad del investigador y fomentarían la extensión de los resultados?
 5. Con respecto al problema 3, ¿qué estrategias de investigación podrían mejorar la validez del diseño, disminuir la parcialidad del investigador y fomentar la ampliación de la información?

Estrategias cualitativas



PALABRAS CLAVE

multimétodo	entrevista-conversación informal
estrategias	guía para la entrevista
problemas previstos	patrones de entrevistas abiertas-cerradas
énfasis holístico	entrevista con informador esencial
trazado del mapa de campo	entrevistas sobre la carrera
cometido de la investigación	profesional y sobre la vida
observación del participante	entrevistas de casos atípicos
estancia en el campo	entrevistas fenomenológicas
notas de campo	tanteos
recogida inmediata de datos	elaboraciones de entrevistas
entrevistas en profundidad	selecciones de artefactos/enseres

Muchos de nosotros estamos acostumbrados a utilizar las conversaciones y las observaciones para adquirir información útil sobre nuestros estudiantes, colegas, organizaciones en las que trabajamos e, incluso, sobre nuestras familias y amigos. Interpretamos nuestras observaciones y compartimos con los demás breves relatos sobre lo que hemos aprendido. En la investigación cualitativa, los distintos tipos de conversaciones y las observaciones constituyen la estrategia básica de recogida de datos.

La investigación cualitativa es investigación *interactiva cara a cara*, lo que requiere un espacio de tiempo relativamente extenso para observar sistemáticamente, entrevistar y registrar hechos cuando éstos tienen lugar de forma espontánea. Las estrategias de la recogida de datos se centran en lo que el fenómeno significa para los participantes. A pesar de la variación considerable entre los estudios cualitativos, las estrategias metodológicas comunes distinguen este estilo de investigación: la observación participante, las entrevistas en profundidad y la selección de artefactos. La mayor parte de los estudios cualitativos son preparatorios y explicativos para comprender los distintos puntos de vista que la gente tiene del mundo. Otros estudios más tradicionales emplean las mismas técnicas pero con un objetivo liberador. Este capítulo presenta la entrada al campo, la selección de un cometido apropiado para la investigación, las estrategias de investigación interactivas y las pautas para evaluar las técnicas cualitativas.

ESTRATEGIAS MULTIMÉTODO

La mayoría de las investigaciones cualitativas dependen de estrategias multimétodo para recopilar datos. Multimétodo es el empleo de varias estrategias con

el objetivo de recopilar y corroborar los datos obtenidos gracias a cada una de las estrategias y /o formas de confirmar datos dentro de una única estrategia de recogida de datos. Los investigadores cualitativos, llamados trabajadores de campo, asumen que cualquier dato puede ser corroborado durante la recogida de datos. Por ejemplo, los datos obtenidos a partir de la observación directa de clases de primaria pueden ser confirmados a través de eventuales entrevistas con profesores y alumnos y en entrevistas más formales con el supervisor principal. Un estudio de los métodos y de la administración de una escuela corroboraría los datos mediante entrevistas, observaciones de las reuniones y las audiciones, reportajes de la prensa y acontecimientos, informes de grupos de interés y documentos del Departamento de Educación.

Los investigadores cualitativos tienen en cuenta la observación participante, las entrevistas, el análisis de los artefactos, la observación del campo y las técnicas suplementarias como estrategias. Los investigadores prefieren las estrategias antes que los procedimientos rígidos paso a paso. Las estrategias son técnicas que sirven para recoger muestras y datos que son perfeccionadas constantemente durante todo el proceso de la adquisición de datos para aumentar la validez de éstos. El empleo de estrategias permite que surja el diseño o que éste vaya tomando forma, parcialmente, conforme los datos son formulados y analizados.

Normalmente, los investigadores escogen una estrategia de recogida de datos básica como la observación participante o las entrevistas y utiliza las otras estrategias para verificar los descubrimientos más importantes antes de abandonar definitivamente el campo. Más adelante se explicará cómo se emplea cada estrategia. Los investigadores determinarán qué métodos son los más prácticos, factibles, eficientes y éticos para realizar la recogida de datos conforme avanza la investigación. Si el objetivo de la investigación es la descripción de los procesos, de los acontecimientos o de los conceptos; entonces, la recogida de datos y las actividades para reunir muestras son sólo un medio para conseguir datos válidos. El uso de las estrategias multimétodo, también, aumenta la credibilidad del estudio. Hablaremos de las siguientes estrategias en este capítulo: la observación participante, las entrevistas en profundidad, la selección de artefactos, la observación de campo y las estrategias suplementarias.

PROBLEMAS PREVISTOS Y REFORMULACIONES

Los investigadores cualitativos empiezan con los **problemas previstos**, problemas previos a la investigación que serán reformulados en el campo durante la recogida de datos. Los problemas previstos son, normalmente, preguntas de investigación generales sobre los participantes (tiempo, lugar, acontecimientos); *qué ocurre, por qué pasa esto y de qué manera tiene lugar*. Tales problemas

pueden centrarse en la estructura y en los procesos de diferentes escenarios sociales y experiencias. Aquí tenemos un ejemplo:

Para describir la tarea personal del director, es necesario describir y analizar tres puntos. En primer lugar, ¿qué comportamientos del profesor consideraron insatisfactorios los directores? En segundo lugar, una vez identificados los comportamientos insatisfactorios, ¿qué hicieron los directores para resolver dichos problemas? Y en tercer lugar, ¿qué factores influyeron en la toma de decisión de los directores durante el proceso de identificación-resolución? (Luck, 1985, p. 4).

Los problemas de la investigación cualitativa provienen de varias fuentes posibles: acontecimientos comunes que se repiten de forma periódica en la educación o en las experiencias personales; ideologías y filosofías, teorías, investigación previa y problemas e ideas señalados por terceros en la preparación de la experiencia. En otras palabras, los problemas empíricos se encuentran dispersos bajo distintas formas y, en la mayoría de los casos, sólo se necesita reconocerlos por sus posibilidades. Existe un sentimiento casi intuitivo de los problemas en la manera de formular preguntas como: ¿me pregunto qué pasará ahora que...?, ¿qué significa realmente este acontecimiento para los participantes? o ¿cómo se las ingeniarán para hacer esto? Reconocer un posible problema de investigación es una destreza etnográfica. Los investigadores poseen una curiosidad extrema y se esfuerzan por desvelar los porqué y los cómo de lo que observan o experimentan. Los investigadores cualitativos estudian la teoría y la investigación previa tanto como el resto de los investigadores, pero aquéllos abandonan, adrede, a un lado este conocimiento hasta que su experiencia en el campo revela su importancia.

El enunciado de los problemas previstos indica que el investigador ha decidido provisionalmente cuál será el objetivo de la investigación, las preguntas y el objetivo de las estrategias de la recogida de datos. La mayor parte de los problemas previstos tiene una importancia descriptiva.

Reformulación de las preguntas. El descubrimiento de las «intenciones» del participante requiere la reformulación de los problemas en el campo. Malinowski, un célebre antropólogo, destacó la habilidad esencial del investigador para refundir las ideas iniciales durante el trabajo de campo cuando declaró:

Si un hombre que se propone hacer una expedición, decidido a demostrar ciertas hipótesis, es incapaz de cambiar sus puntos de vista constantemente y de desecharlos sin ningún reparo bajo la presión de la evidencia, inútil decir que su trabajo no valdrá la pena. Pero, cuantos más problemas aporte al campo, cuanto más acostumbrado esté a analizar los hechos en relación con la teoría, mejor equipado estará para el trabajo... (1922 [Reimpresión, 1961], p. 8-9)

En un estudio exploratorio (Schumacher, 1975) sobre la entrada en vigor del plan de estudios (uso extendido) en un estado, por ejemplo, los problemas pre-

vistos fueron expresados en cuanto a los acontecimientos secuenciales planeados para conducir a la adopción del programa de estudios en las escuelas de una región. A pesar de la drástica modificación de los planes y de los acontecimientos programados para el año, la entrada en vigor del programa de estudios tuvo lugar. Los iniciales problemas previstos fueron reformulados para centrarse más en la relación coste-beneficio y para llegar a un acuerdo entre el Departamento estatal de Educación y los sistemas de la escuela local.

Los problemas previstos *no* son ideas preconcebidas sino un conocimiento adquirido en el trabajo sobre los hechos, las consecuencias, los conceptos y las teorías que guían las decisiones durante la recogida de datos. La investigación cualitativa representa un acercamiento ecléctico hacia el uso de teorías en un estudio. En general, las teorías pueden influir en las preguntas de la investigación de dos maneras: creando preguntas de investigación o creando estructuras conceptuales sobre las preguntas iniciales. Los alumnos que ya han superado ciertas asignaturas, normalmente, plantean preguntas que reflejan los conceptos y las suposiciones de esta asignatura. Si entramos en el campo con una perspectiva sociológica, psicológica, antropológica, política o crítica, es importante que lo indiquemos explícitamente. Los investigadores que entran en el campo con varias estructuras conceptuales, pueden reconocer más fácilmente los acontecimientos para sacar a relucir los significados latentes.

El objetivo de las estrategias de la recogida de datos. Los problemas previstos indican el objetivo de la recogida de datos. Esto es muy importante a la hora de escoger el/los lugar/es para la observación participante y, también, para la selección de los participantes en las entrevistas en profundidad. No obstante, los problemas previstos *no* limitan la investigación, porque muchas preguntas de la investigación evolucionan durante el largo campo de trabajo. El extracto 11.1 ilustra la selección de una escuela de primaria de zona urbana para identificar las prácticas de educación específicas de un modelo de integración con éxito, pero la reformulación de las preguntas iniciales conduce a centrarse en la cultura de la totalidad de la escuela que era partidaria de la integración.

ENTRADA EN EL CAMPO

El análisis meticuloso de los problemas previstos establecerá criterios sobre el sitio y/o escenario social para la observación participante. Un análisis similar nos indicará los perfiles de los individuos que deben ser seleccionados para las entrevistas en profundidad. Trataremos cada una de estas estrategias para introducirnos en el campo. En las negociaciones que se llevan a cabo con el fin de obtener permiso oficial para realizar estudios cualitativos, el investigador también hace explícito el cometido de la investigación para la recogida de datos.

EXTRACTO 11.1**PROBLEMAS PREVISTOS REFORMULADOS**

La integración de los estudiantes con discapacidades es uno de los mayores movimientos de reforma escolar de nuestro siglo... los investigadores que trabajan en dicha integración han centrado mucha más atención en identificar las prácticas educativas específicas que contribuyen a una integración con éxito (Hunt y Goetz, 1997) que en explorar la conexión entre la integración y la cultura de la escuela.

Este estudio empezó de un modo similar, teniendo por objetivo la identificación de las prácticas educativas específicas que contribuyen al éxito de un programa modelo de integración en una escuela de primaria de zona urbana, la Escuela Primaria Connolly (seudónimo) en una comunidad de diversidad cultural. Durante el año en que duró el proceso de observación participante, de entrevistas y de revisión de documentos, dedujimos, a partir de nuestros datos, que las prácticas educativas eran uno de los aspectos de una cultura escolar más amplia que estaba completamente a favor de la integración. Esperamos que, gracias a la identificación de las características de esta «cultura escolar integradora», sirvamos de ayuda a los futuros esfuerzos por llevar a cabo programas de integración.

Fuente: Zollers, N. J. , Ramanathan, A. K. y Yu, M. (1999), «The relationship between school cultura and inclusión: How an inclusive cultura supports inclusive education», *International Journal of Qualitative Studies in Education*, 12 (2), 157-174.

Selección del lugar y trazado del mapa de campo

Escoger un lugar es un proceso de negociación para obtener libertad de acceso a un escenario que resulte idóneo para los problemas de investigación y viable para los investigadores según su disponibilidad de tiempo, de movilidad y de cualidades. El investigador de campo, normalmente, obtiene información de antemano a través de diferentes medios. Como información útil podríamos citar las identidades, las alineaciones de poder y los intereses de los principales participantes; la historia general, las costumbres, el sistema social y las actividades del lugar. La información que concierne al lugar y su potencial idoneidad se puede obtener a partir de fuentes muy diversas: documentos, principales asociaciones actuales e información pública. Mucho depende del buen juicio del investigador, el tiempo que emplee y el tacto que posea para recopilar la información.

Después de que el investigador haya identificado un lugar posible, contacta con la persona que pueda conceder el permiso para realizar la investigación. Algunos investigadores llevan a cabo un contacto formal tras la confirmación informal de que la propuesta de investigación será estudiada. La mayoría escriben un breve informe en el que especifican el lugar, los participantes y las actividades, la duración total del estudio y el objetivo de la investigación. El documento,

además, proporciona información sobre el investigador, el patrocinador o la organización, y el empleo general de los datos, incluyendo la protección de los derechos de las personas. El acceso al lugar y a las personas resulta crucial en este momento. Schatzman y Strauss sugieren que «cualquier restricción inicialmente aceptada por el investigador debería ser considerada como renegociable más tarde, en tiempos más propicios» (1973, p. 29). Una autorización formal es esencial para la moralidad de la investigación y para proceder a la entrada en el campo y establecer el cometido de la investigación. Una vez que la autorización ha sido otorgada, el investigador toma el mando del lugar para trazar el mapa de campo, dirigir el muestreo y establecer el cometido de la investigación.

La investigación cualitativa describe e interpreta cualquier subconjunto de datos contextualmente limitados dentro del contexto más amplio del lugar. Esta característica se refiere, a menudo, a un énfasis holístico: los subcasos de datos son relativos al contexto total del fenómeno estudiado. Aunque todo esto ocurra dentro de un marco determinado, es una potencial fuente de datos, los observadores participantes *no pueden y no necesitan observarlo todo*, pero sí pueden obtener suficientes datos para un énfasis holístico. Utilizando estrategias de trazado de mapa, el investigador adquiere una visión de la «totalidad» y parte de una base mejor para llevar a cabo el muestreo de casos con abundante información.

Trazado del mapa de campo. Conseguir la entrada en el campo requiere establecer buenas relaciones con todos los individuos del lugar de investigación. El permiso de investigación se concede sin ninguna garantía de que los participantes se comporten de manera natural ante un visitante que toma notas o de que los participantes compartan sus impresiones, sus pensamientos y sus sentimientos con el observador. Las cualidades del investigador se reflejan en cómo los participantes le reconocen como alguien interesado y respetuoso; un observador que no juzga sino que se mantiene reservado, o si lo ven como alguien maleducado, destructivo, un observador crítico en el que no se puede confiar. El investigador debe ganarse la confianza y la confidencialidad de los participantes, constantemente, a lo largo de todo el periodo de recogida de datos (Schumacher, 1984a). Los participantes deben poder compartir sus impresiones, sentimientos y pensamientos en cualquier momento con el etnógrafo.

El trazado del mapa de campo es ir adquiriendo datos de las relaciones sociales, espaciales y temporales del lugar con el fin de adquirir una visión total del contexto. En un mapa *social* se anota la cantidad y tipos de personas, la estructura organizativa y las actividades a las que se dedica la gente. En un mapa *espacial* se anotan las ubicaciones, facilidades y servicios especializados proporcionados. En un mapa *temporal* se describe el ritmo de organización de la vida, los horarios y las costumbres.

Muestras del lugar. Una vez trazado el mapa de campo, los investigadores seleccionan a las personas, las situaciones y los acontecimientos con más

probabilidades de proporcionar datos sobre las preguntas de la investigación. Al principio, el investigador busca informadores con información abundante, grupos, lugares o acontecimientos de los que seleccionar subunidades para un estudio más detallado. No debemos olvidar que el conjunto de muestras es una estrategia para escoger pequeños grupos o individuos de los que se puedan obtener conocimientos e información sobre el tema de interés (ver el capítulo 10). Además, la mayor parte de los investigadores no saben de antemano si posibles casos con abundante información proporcionarán o no datos válidos hasta que haya completado un análisis parcial de datos. Así que, normalmente, se planea el conjunto de muestras al principio y, luego, le siguen estrategias adicionales.

Conforme van apareciendo nuevas preguntas durante la recogida de datos, el observador modifica (a través de más conjuntos de muestras) las horas de observación y los lugares de recogida de datos válidos. El extracto 11.2 ilustra la relación entre informes problemáticos, la selección del lugar, el diseño y el trazado del mapa de campo en un estudio sobre el proceso de admisión de niños con discapacidades psíquicas con el objetivo de considerar su institucionalización.

Los trabajadores de campo consideran que las estrategias de selección y conjunto de muestras, son dinámicas y más adecuadas que los parámetros estadísticos o *a priori* de poblaciones para el diseño de la investigación. Así pues, la recogida de muestras es un proceso llevado a cabo simultáneamente con la recogida de datos. La especificación de la selección de criterios y la estrategia de recogida de muestras escogidas se reunirán en un estudio con el fin de reducir posibles amenazas a la validez del diseño. Además, la persona o grupos que participan en el estudio serán presentados de tal manera que la confidencialidad de los datos quede protegida.

Selección de los entrevistados

La selección de las personas para las entrevistas en profundidad empieza con una descripción de las cualidades o perfil deseado de las personas que poseen conocimiento sobre el tema en cuestión. Una cualidad es que todas las personas hayan vivido una experiencia similar, como por ejemplo, haber dimitido voluntariamente de un puesto administrativo, haber vivido la muerte de un niño o haber participado ya en otro programa.

La localización de posibles entrevistados puede hacerse a través del uso de documentos, de la red de información o por denominación. Algunos ejemplos sobre la utilización de documentos serían las afiliaciones y los registros obligatorios, como la solicitud para un programa o un servicio especial. Cuando cada persona no interviene en situaciones de cara a cara con otros conocidos para tener experiencias similares, el muestreo en cascada o de bola de nieve resulta decisivo (ver el capítulo 10). Los entrevistados también pueden ser escogidos por su reputación como el «mejor» profesor o entrenador.

EXTRACTO 11.2**PROBLEMAS DE INFORME, SELECCIÓN DEL LUGAR, DISEÑO BUSCADO, TRAZADO DEL MAPA DE CAMPO**

Quería saber qué pensaba la gente directamente afectada o involucrada en un proceso de compromiso institucional y precisar sus experiencias...

El diseño del proyecto consistía en seguir a aquellas personas que fueron admitidas en el *Weston Center* y/o solicitaron que se les admitiera o readmitiera en dicho centro entre el 1 de enero y el 31 de diciembre de 1981.

De un total de 17 personas, aquí figuran 13. Una no se incluyó porque no pude ponerme en contacto con la familia para el permiso. Otros dos no se incluyeron porque las familias no querían participar. El cuarto no se incluyó porque en el momento en que me puse en contacto con el padre y éste aceptó formar parte del proyecto, la recogida de datos ya estaba completa...

CONSEGUIR ENTRAR EN EL *WESTON CENTER*

Me acerqué a *Weston Center* en diciembre de 1980. Durante enero y febrero de 1981, su Comité de Investigación revisó mi propuesta y establecieron las líneas directrices conforme a lo que yo tenía que realizar en mi investigación. Se pidió a los padres y al personal que firmaran las hojas de «consentimiento» antes de que me diera tiempo a incluirlos en el estudio. Me dieron acceso completo al centro y me permitieron asistir a todas las reuniones que considerara relevantes para mi proyecto.

Pero, incluso con esta «carta blanca» por parte de la administración, me di cuenta de que mi entrada en el campo dependía de la aceptación por parte del personal de todos los niveles de la organización. Consecuentemente, estuve varias semanas presentándome al personal, hablando del proyecto con ellos y respondiendo a sus preguntas.

En ocasiones, las ideas que surgían de estos encuentros constituirían la prueba instrumental del posterior éxito del estudio. Por ejemplo, se me sugirió que el trabajador social encargado del caso me presentara personalmente a los padres, en lugar de hacer mis preguntas por teléfono o por correo electrónico. Si bien es cierto que no podemos comparar este acercamiento con lo que hubiera ocurrido sin una presentación personal, sí es importante destacar que de entre las dieciséis familias contactadas, tan sólo dos se negaron a participar en el estudio. Además, nadie del personal se opuso a una entrevista.

Fuente: Foster, S. B. (1987), *The Politics of Caring*, Londres: Falmer Presse Ltd.). Reimpreso con permiso de Falmer Press Ltd. y el autor.

Los investigadores examinan a cada uno de los entrevistados potenciales a través de las cualidades o perfiles desarrollados en el estudio. Sólo los individuos que muestran los criterios requeridos serán entrevistados en profundidad.

Cometido de la investigación

Los trabajadores de campo escogen una función dentro de la investigación (relaciones adquiridas y atribuidas al investigador en la recogida interactiva de datos) apropiada al objetivo del estudio. Entre los cometidos posibles están el observador total, el participante total, el observador-participante, el observador-confidente, entrevistador y la doble función de investigador-participante (ver tabla 11.1). Estos cometidos varían en la medida en que la presencia del investigador afecte al sistema social o a las personas bajo estudio. El papel de observador total es el de alguien que está física y psicológicamente ausente. Un ejemplo es el del observador que mira a través de una única ventana. El papel de participante total es similar al de vivir a través de una experiencia, recordándola y escribiendo lo que se siente.

Las actuaciones de observador-participante o entrevistador son los cometidos de la investigación más usados por los formularios de preguntas cualitativas. El trabajador de campo se permite, según la ocasión, adoptar un papel con el único objetivo de realizar la recogida de datos. El observador-participante obtiene permiso de la organización o del grupo cultural; los trabajadores de campo deben es-

TABLA 11.1: COMETIDOS POSIBLES DE LA INVESTIGACIÓN INTERACTIVA

	Descripción	Uso	Método de investigación
Observador total	El investigador está física y psicológicamente ausente.	Inapropiado	Otros tipos de investigación.
Participante total	El investigador vive a través de una experiencia y recopila sentimientos personales.	Inapropiado	Ninguno.
Observador-participante	El investigador crea su papel según el objetivo del estudio.	Cometido típico	Etnografía. Estudio de caso. Algunos estudios críticos.
Observador-confidente	El investigador tiene una posición formal en la organización.	Inapropiado	Muy inusual (depende del tema)
Entrevistador	El investigador establece un papel con cada persona entrevistada.	Cometido típico	Estudios fenomenológicos. Teoría inductiva. Algunos estudios críticos.
Investigador-participante	El investigador establece un papel doble según el objetivo del estudio.	Es muy difícil realizar ambos cometidos simultáneamente.	Estudios críticos: participación en la investigación, interrogatorios.

tablecer cuál será su cometido para la investigación con cada persona o grupo seleccionado para el estudio. Por el contrario, el entrevistador empieza a desempeñar el papel en el primer contacto con la persona cuando le solicita una cita y le explica los objetivos y el carácter confidencial. El entrevistado elige el momento y el lugar de la entrevista. En estudios fenomenológicos, los entrevistadores pueden reunirse tres o cuatro veces con cada persona para conseguir comprender la experiencia vivida.

Otras actuaciones discutibles son las de observador-confidente e investigador-participante. El observador-confidente es una persona que ya desempeña una función en el lugar en el que se propone realizar el estudio. La función (puede ser, director, asesor, profesor) existe, independientemente de que el estudio se realice o no. Incluso, los estudios sobre temas muy delicados, que probablemente no podrían haberse llevado a cabo de otro modo, tienen dificultad de ser aceptados como una investigación creíble. Algunos estudios críticos emplean el doble cometido de investigador-participante. Este papel se desempeña simultáneamente y requiere que el investigador sea consciente constantemente sobre de quién es la voz que está siendo recogida como dato. Porque los datos recogen las reflexiones del investigador durante la experiencia igual que las de los participantes «reales». El doble cometido de investigador debe ser sumamente estudiado para así asegurarse de que la voz que quedará recogida en el estudio, será la correcta.

El cometido de la investigación está constituido, en realidad, por muchos papeles, a medida que el trabajador de campo va adquiriendo fluidez verbal con los participantes, interactúa para obtener datos, establece relaciones sociales y cambia de un papel a otro según el grupo o persona. Normalmente, siempre es necesario cierto grado de participación para que el desconocido llegue a confiar y para conseguir reciprocidad. Además, la función desempeñada dentro de la investigación puede variar según el grado de interacción e intensidad. Por ejemplo, en muchos estudios etnográficos, estudios de caso y teoría inductiva del acercamiento, las interacciones son bastante satisfactorias pero el investigador es menos dado a la recogida de datos. En estudios fenomenológicos, la interacción es más cercana y personal. A diferencia de los recursos mecánicos, los trabajadores de campo son capaces de añadir otras preguntas, revisar todos los puntos y profundizar en el análisis del fenómeno.

Los datos válidos se obtienen cuando los acontecimientos se revelan espontáneamente y los participantes actúan de la manera en que suelen hacerlo ante la presencia del investigador. Como el papel en la investigación afecta el tipo de datos recogidos, los papeles (tanto el principal como los secundarios) adoptados durante la recogida de datos, quedan recopilados en el estudio. El extracto 11.3 es un ejemplo de los papeles y los tipos de actividades de recogida de datos llevados a cabo en un estudio etnográfico.

EXTRACTO 11.3**PAPEL PARA LA INVESTIGACIÓN**

En nuestro estudio etnográfico, que se prolongó durante un año, empleamos la observación de los participantes, las entrevistas formales e imprevistas y los análisis de documentos... La observación de los participantes en el medio escolar nos permitió obtener datos modesta y sistemáticamente e interactuar socialmente con las personas que nos aportaban la información... Visitábamos la escuela por lo menos una vez a la semana. En los días de nuestras visitas, participábamos en la escuela, visitando clases y hablando con padres y personal en la entrada. Observábamos a los profesores, a los alumnos y al personal a lo largo del curso normal de cualquier día de escuela y durante los acontecimientos especiales.

Fuente: Zollers, N. J., Ramanathan, A. K., y Yu, M. (1999), «The relationship between school cultura and inclusión: How an inclusive cultura supports inclusive education», *International Journal of Qualitative Studies in Education*, 12 (2), 157-174.

LA OBSERVACIÓN PARTICIPANTE

La observación participante es realmente una combinación de estrategias particulares de recogida de datos: participación limitada, observación de campo, entrevistas y la recogida de artefactos. La participación limitada es necesaria para obtener una aceptación de la presencia del investigador, incluso si se trata de alguien discreto. La observación de campo es la técnica del investigador que le permite observar directamente y tomar notas sin interacciones. Las entrevistas pueden tener lugar en forma de conversaciones casuales tras una vivencia con otros (informal) o pueden ser más formales y con una única persona. Los documentos y los artefactos se utilizan cuando están disponibles. Normalmente, el investigador utiliza múltiples estrategias para corroborar los datos. El extracto 11.4 ilustra estrategias multimétodo.

Las decisiones que conciernen a las estrategias de recogida de datos son normalmente revisadas tras la selección del lugar, la entrada en el campo y el trazado inicial del mapa de campo. Los planes iniciales se pulen y revisan. La información que, en un principio, se había planeado que se obtendría a través de la observación puede conseguirse a través de entrevistas; es posible que los descubrimientos preliminares procedentes de los datos de las entrevistas, tengan que comprobarse con la ayuda de cuestionarios elaborados por el investigador o con herramientas de análisis. El análisis de los documentos o de los acontecimientos inesperados puede abrir nuevas direcciones para la observación y las entrevistas. Elegir las estrategias de recogida de datos es un proceso en el que se debe escoger entre varias alternativas para la recogida y corroboración de datos y en el que debemos modificar nuestras decisiones para captar la «realidad» del fenómeno.

La observación participante *in situ*. El requisito más elemental de la metodología es la **permanencia en el campo**, en el que el investigador está presente durante un tiempo más prolongado. El trabajo de campo, a menudo, es considerado como un esforzado trabajo de búsqueda. Muchos estudios de campo se centran en procesos sobre el tiempo y la modificación de notas. En el extracto 11.5: ejemplo A, un estudio sobre el poder y el cuidado, el etnógrafo pasa un día completo cada semana observando la clase de Pam (la profesora) a lo largo de todo un curso académico. Los investigadores cualitativos normalmente dejan constancia del número de días y de horas de permanencia en el campo como en el extracto 11.5: ejemplo B.

Recogida prolongada de datos. La recogida de datos continúa hasta el fin lógico de un acontecimiento o hasta que la situación cambie tan radicalmente que el escenario no sea relevante para continuar la investigación. El límite natural para la recogida de datos puede ser un periodo de tres semanas para un programa artístico en una escuela de verano. El límite natural para un estudio de instrucción en enfermería clínica es de diez semanas. Cuando la situación examinada ya no resulta relevante para la investigación, la permanencia en el campo finaliza. La recogida de datos puede terminar con la inesperada promoción o renuncia de una persona decisiva, circunstancia que, evidentemente, supondría su desaparición en el estudio.

EXTRACTO 11.4

ESTRATEGIAS MULTIMÉTODO. RECOGIDA DE DATOS

Documentos archivados

En cuanto el proyecto fue aprobado, empecé la recogida de datos. Primero, revisé los historiales médicos de las personas incluidas en el estudio, así como los documentos generales oficiales que describían el *Weston Center* y sus actividades. Los documentos (textos formales de identificación y descripción) constituyen una gran ayuda para comprender mejor a los individuos (Biklen, 1978). El objetivo de esta revisión de los documentos era adquirir esa perspectiva.

Observación

En segundo lugar, pasé tiempo como observador en el *Weston Center*. Acudía a reuniones que eran importantes para la aceptación, la evaluación y el establecimiento del proceso en el centro. En varios casos, tales observaciones se extendían a aulas de cuidados básicos y a las clases del centro.

Notas de campo

Al final de cada observación, escribía las notas de campo describiendo con detalle los temas tratados en las actividades, etc.

Otras observaciones

Sin embargo, no fue posible recoger notas de campo a lo largo de las observaciones de otros medios, como la organización de aulas de cuidados básicos y de las oficinas del departamento.

Resumen de observaciones

Con el fin de asegurar la exactitud y detalle de dichas notas, debía emplear varias horas para registrar todos los datos tras abandonar el campo.

(*continúa*)

(continuación)

Los dibujos en las notas de campo
Conversaciones informales como entrevistas

Los dibujos fueron un elemento importante en las notas de campo...

Además de estas observaciones, entrevisté a participantes en el proceso de organización, incluyendo personal del centro, otras agencias de servicio social y miembros de las familias de los alumnos. Esas entrevistas se desarrollaron más como conversaciones que como entrevistas formales y las realizamos en una amplia variedad de lugares, incluyendo oficinas, salones, coches y vestíbulos. El objetivo de las entrevistas era permitir a aquéllos envueltos en la organización del proceso, describir con detalle sus percepciones e interpretaciones de la experiencia.

Guía de acercamiento a través de las entrevistas; redactar y elaborar

Llevaba una grabadora a las entrevistas programadas y, con el permiso del entrevistado, registraba nuestra conversación. No obstante, también grababa las notas de campo de observación para cada entrevista, describiendo el lugar en el que la entrevista tenía lugar, la persona y sus acciones. A menudo, dichos detalles resultarían ser tan importantes como la misma entrevista.

Otras conversaciones

A veces, las conversaciones al principio o al final de los encuentros proporcionaban una gran cantidad de datos. Una vez, una trabajadora social a cuyo paciente se le había negado la admisión, me paró en el recibidor para explicarme cuál era su interpretación de la política que rodeaba dicha decisión. En casos como éste, normalmente, yo empleaba mucho tiempo en escribir notas en mi coche antes de volver a casa y, así, poder obtener una reproducción de la conversación lo más esmerada y detallada posible.

Foster, S. B. (1987), *The Politics of Caring*. Londres: Falmer Presse Ltd.). Reimpreso con permiso de Falmer Press Ltd. y el autor.

EXTRACTO 11.5

RECOGIDA PROLONGADA DE DATOS

EJEMPLO A

Creo que deberían saber que el poder de Pam (la profesora) se extendía mucho más allá de la clase. Ella era una de los cabecillas del establecimiento, era venerada por padres blancos y de origen africano, y era la profesora que asumía el cambio de escuela siempre que el director no estaba presente. Tenía la reputación de ser la profesora más eficaz del establecimiento... Normalmente, yo pasaba un día entero a la semana observando su clase durante el curso escolar 1989-90. Había pasado de ser un profesor titulado en Educación por la universidad, a ser su «estudiante mayor».

(continúa)

(continuación)

EJEMPLO B**TABLA 1: Tiempo empleado en la recogida de datos en horas**

	Entrevistas								
	Formales (grabadas)			Informales		Observación		Revisión de documentos	
	Profesorado/ personal	Padres	Sr. Knight	Sr. Knight	Personal y padres	Campus	Fuera del campus	Artículos/ evaluación	Composiciones de estudiantes
Frecuencia	9	4	5	16	5	16	1	20+	38
Horas totales	6	3	5	8	8	96	.5		

Fuente: *Example A*—From Nolbit, G. W. (1993). Power and caring. *American Educational Research Journal*, 30(1), 23–38; *Example B*—From Zollers, N. J., Ramanathan, A. K., y Yu, M. (1999). The relationship between school culture and inclusion: How an inclusive culture supports inclusive education. *International Journal of Qualitative Studies in Education*, 12(2), 157–174.

Observación y escucha intensas. La observación participante permite al investigador conocer su percepción de los acontecimientos y de los procesos expresados a través de sus acciones y expresados como sentimientos, pensamientos y opiniones. Estas percepciones o interpretaciones pueden adoptar tres formas: verbal, no verbal y conocimiento sobreentendido. Es crucial que el investigador adquiera las pautas lingüísticas particulares y las variaciones del lenguaje de los individuos observados para recoger e interactuar con ellos. La observación de campo es un proceso activo que incluye indicaciones *no verbales* (expresiones faciales, gestos, tono de la voz, movimientos corporales y otras interacciones sociales no verbales que dan a entender los significados más sutiles del lenguaje). El *conocimiento sobreentendido* es personal e intuitivo, por lo que al sujeto le resulta difícil o imposible expresarlo con palabras; en lugar de ello, la persona demuestra este conocimiento a través de sus acciones o mediante la creación de objetos. Muchas culturas reciben el calificativo de expresivas porque los significados se transmiten más de manera no verbal que verbal. Las historias de los participantes, sus anécdotas y mitos —que podemos encontrar en las charlas diarias en la sala de profesores o entre grupos de estudiantes en las entradas— indican el contenido de sus mundos y cómo lo perciben.

La escucha también es un trabajo necesario; los investigadores escuchan con todos sus sentidos. Escuchar implica ser capaz de ponerse en la piel del otro, ser capaz de ver el mundo como el participante. El trabajador de campo presta mucha atención a los *es* y a los *porqué*. El *es* informa acerca de las percepciones de las cosas, de la gente, de los acontecimientos y de los procesos que la persona considera reales o irreales. El *porqué* informa sobre el *porqué* y las razones de las opiniones, los pensamientos, los sentimientos y los valores, en resumen, la lógica sobre el contenido de las impresiones de una persona. La escucha activa

exige que el investigador deje a un lado sus propios pensamientos y coloque en primer lugar los de los participantes.

Las observaciones de campo ratificadoras. Aunque el trabajador de campo no interviene, sí adopta diferentes puntos de vista sobre los acontecimientos de los participantes para conseguir exactitud y confirmación de sus datos. Gracias a la observación de los participantes en muchos contextos, los investigadores consiguen obtener datos que serían «prácticamente imposibles de obtener a través de otras técnicas, y tienen acceso a un tipo de información única» (Wilson, 1977, p. 256). Los etnógrafos pueden, por ejemplo, corroborar lo que un participante contesta como respuesta a un comentario o a una pregunta, a otras personas, en diferentes situaciones o en distintos momentos; lo que el participante hace en realidad; lo que el participante deja entrever con la comunicación no verbal, como el tono de la voz, los movimientos del cuerpo, y lo que percibe de lo que los demás sienten, expresan o hacen sobre una actividad.

La clave de recopilar y corroborar diferentes percepciones reside en obtener datos de múltiples fuentes de datos –diferentes personas en diferentes contextos y momentos distintos–. Las múltiples fuentes de datos pueden tratarse mejor en un estudio si elaboramos una lista de las estrategias, los participantes, las situaciones o las organizaciones. La tabla 11.2 ilustra las múltiples fuentes de datos en un estudio de desarrollo de los planes de estudio.

En la recogida para corroborar datos, los investigadores presentan, a menudo, discrepancias entre lo que la gente dice y lo que la gente hace. De este modo, la observación participante permite comprobar si lo que dice es lo mismo que lo que hace, si esto último coincide con lo que el investigador cree que aquél está haciendo basándose en los datos.

TABLA 11.2: MÚLTIPLES FUENTES DE DATOS

Métodos múltiples	Observación. Conversaciones casuales. Entrevistas dirigidas. Documentos: ofertas de proyecto, documentos sobre programas de estudios sociales, inspecciones del estado, idea general de las lecciones, etc.
Participantes múltiples	Oficiales del departamento estatal, división escolar del personal de la oficina central, especialista en plan de estudios, director y profesores.
Situaciones múltiples	Reuniones del consejo, taller de verano, reuniones del profesorado, el centro de planes de estudio de la universidad, el departamento de conferencias.

Fuente: adaptado de Schumacher, S. (1979), *Ethnographic inquiry: Theory and application in educational research and evaluation*. (Richmond, VA: Virginia Commonwealth University) (ERIC Document Reproduction Service No. ED 184 108), p. 11.

Observaciones de campo principales. Puesto que el escenario social interactivo es demasiado complejo y sutil para observarlo y captarlo todo, el investigador *no recopila* todo lo que ocurre. Los trabajadores de campo cuentan con la larga estancia en el campo para desarrollar el estudio y, de esta manera, poder, más adelante, decidir lo que debe incluirse o excluirse. Los investigadores observan y recopilan los fenómenos más destacados de los problemas previstos, sus estructuras conceptuales más extensas y las características contextuales de las interacciones. Estos elementos explican la diversidad de notas de campo citadas en etnografías a pesar de lo común de la metodología.

¿Qué observan los trabajadores de campo? La mayoría recopilan detalles descriptivos sobre *quién, qué, dónde, cómo y por qué* una actividad o acontecimiento social ocurrió. Esta información puede ser utilizada para obtener información más sutil (ver la tabla 11.3). Aunque nadie formula todas estas preguntas al mismo tiempo durante el estudio de un acontecimiento, la estructura nos muestra áreas de observación más amplias.

TABLA 11.3: CUADRÍCULA SOBRE LA OBSERVACIÓN DEL PARTICIPANTE

Observación	Descripción
1. ¿Quién hay en el grupo o en la escena?	¿Cuántas personas hay presentes? ¿cómo son o cuáles son sus identidades? ¿Cómo se acepta su ingreso en el grupo o escena?
2. ¿Qué ocurre aquí? a. ¿Qué comportamientos son repetitivos e irregulares? b. ¿Cómo se comportan las personas del grupo entre ellas? c. ¿Cuál es el contenido de sus conversaciones?	¿Qué se dicen unos a otros y qué hacen? ¿En qué actividades o costumbres participa la gente? ¿Cómo se organizan, se explican y justifican las actividades? ¿Cómo se organiza la gente o cómo se cuentan sus vivencias los unos a los otros? ¿Qué papeles resultan evidentes? ¿Quién decide por quién? ¿Qué temas se tratan a menudo y cuáles nunca? ¿Qué lenguajes utilizan para la comunicación verbal y no verbal? ¿Qué opiniones quedan reflejadas a través de sus conversaciones? ¿Qué formatos y procesos siguen las conversaciones? ¿Quién habla y quién escucha?
3. ¿Dónde se sitúa el grupo o la escena?	¿Qué lugares físicos constituyen sus contextos? ¿Qué fuentes naturales y tecnologías se crean o se emplean? ¿Cómo se ha colocado el grupo y cómo utiliza el espacio y los objetos físicos? ¿Qué signos, sonidos, olores, gustos y sentimientos se perciben en los contextos del grupo?

(continúa)

(continuación)

4. ¿Cuándo se reúne el grupo e interactúa?	¿Con qué frecuencia se realizan las reuniones y cuánto duran? ¿Cómo se usa y se distribuye el tiempo? ¿Cómo ven los participantes su pasado y su futuro?
5. ¿Cómo se interrelacionan los elementos identificados (desde la perspectiva de los participantes o del investigador)?	¿Cómo se mantiene la estabilidad? ¿Cómo se produce un cambio? ¿Qué reglas y normas gobiernan esta organización social? ¿Qué se cuenta de este grupo a otros grupos, organizaciones o instituciones?
6. ¿Por qué el grupo actúa como lo hace?	¿Qué significado atribuyen los participantes a lo que hacen? ¿Qué símbolos, tradiciones, valores y puntos de vista del mundo pueden encontrarse en el grupo?

Las notas de campo y su registro inmediato. Los datos se recopilan en forma de **notas de campo**, observaciones de lo que acontece mientras el investigador está en el campo. Las notas de campo aparecen fechadas y el contexto se identifica. A menudo, las notas están repletas de abreviaturas características que resultan difíciles de leer para terceros a no ser que se redacten. Los investigadores recopilan campos descriptivos detallados y muy precisos. Aquí abajo encontramos algunos ejemplos hipotéticos de notas imprecisas contrastadas con las notas de campo que hoy día se recogen en un estudio sobre lectores adultos principiantes (Boraks y Schumacher, 1981, p. 76, 86).

Notas imprecisas	Notas descriptivas
1. Mientras lee, Bea confunde las palabras «would» por «wood»	(Observación: el esfuerzo de Bea queda demostrado por su respiración entrecortada e, incluso, por la sudoración durante la lectura. No hay duda de que se esfuerza al máximo.) A veces, cuando lee: <i>Would you believe I would not do it again</i> , Bea pronuncia <i>would</i> como <i>wood</i> . El profesor dice <i>would</i> . Bea intenta volver a leer y luego dice, <i>Oh, I missed the point, would he</i> .
2. June vuelve a contar pequeñas partes de una historia y sólo las partes que se cuenta a ella misma. Elabora estas partes.	June es una mujer alta, delgada y habladora. Un miembro del personal había dicho que se la consideraba retrasada mental y con problemas emocionales. Cuando se le pidió que contara una historia de un texto indicado, se le incitó a hablar. Por ejemplo, su historia sobre un accidente: June: «Mi pelo estaba ardiendo». Profesor: «¿Cómo se prendió fuego? ¿Puedes contarme más detalles?». June: «Estaba fumando y encendí mi mechero. Lo tenía cerca de mi oreja y la llama llegó a mi pelo, así que tenía toda la cabeza en llamas». Profesor: «¿Puedes contarme algo más?». June: «Me dijeron que parecía un árbol de navidad con las luces encendidas».

Los registros inmediatos, escritos inmediatamente después de haber abandonado el lugar, resumen las principales interacciones y escenas observadas y, lo que es más importante, valoran la calidad de los datos, sugieren preguntas y posibles interpretaciones. Tanto en las notas de campo como en los registros inmediatos, las posibles interpretaciones se separan de las observaciones actuales (datos). En ocasiones, estas percepciones pueden crear asociaciones libres: citar analogías, usar metáforas y símiles o anotar teorías y bibliografía que pueda resultar útil en análisis de datos futuros.

Normalmente, una parte de los registros realizados de forma inmediata es una exigencia crítica del propio investigador ante posibles prejuicios. A menudo denominada como «investigación de sí mismo» en el estudio, el investigador se vale de fuentes biográficas de subjetividad y de estrategias para minimizar la influencia en la recogida de datos y un análisis (ver extracto 11.6)

EXTRACTO 11.6

NOTAS INMEDIATAS: INVESTIGACIÓN DE SÍ MISMO

Como muchos otros investigadores cualitativos, nosotros consideramos el acto de la investigación como un acto en el que es difícil realizar una valoración libre... Consideramos nuestra subjetividad como «un vestido que no podemos quitarnos» (Peshkin, 1988:17). En lugar de intentar, inútilmente, eliminar esta subjetividad, Horvat adopta un «control sistemático y formal de sí mismo» durante todo el periodo de la recogida de datos, lo que le permite dominar su propia subjetividad. Para conseguirlo, escribe una nota con sus propias opiniones, comparte manuscritos de datos analizados con los participantes en el estudio y comenta los temas que se van presentando con colegas familiarizados con el proyecto.

Además, creemos que nuestro pasado es lo que da forma a nuestras actuaciones como investigadores. Horvat es una mujer blanca casada, de treinta y pocos que creció en California en una pequeña ciudad situada no muy lejos de donde se realizó esta investigación. Su privilegiado pasado es similar al de dos o tres de los estudiantes más privilegiados de Hadley, económicamente hablando, del estudio,...

Ella asistió a una escuela privada parecida a Hadley. El segundo autor... es un hombre filipino que fue educado por padres inmigrantes... Social y económicamente, su educación es bastante distinta de la de la mayoría de los participantes en el estudio... Como colaboradores de investigación, aprovechamos nuestros diferentes pasados para llamarlos la atención recíprocamente sobre posibles prejuicios.

Fuente: Horvat, E. M. y Antonio, A. (1999), «Hey, those shoes are out of uniform: African American girls in an elite high school and the importance of habitus», *Anthropology & Educational Quarterly*, 30(3), 317-342.

ENTREVISTAS EN PROFUNDIDAD

Las **entrevistas en profundidad** son preguntas con respuesta abierta para obtener datos sobre *los significados* del participante: cómo conciben su mundo los individuos y cómo explican o «dan sentido» a los acontecimientos importantes de sus vidas. Las entrevistas pueden ser la principal estrategia de recogida de datos o una ampliación natural de estrategias de observación. Las entrevistas de campo varían por el formato, las aplicaciones especializadas, el contenido de las preguntas, el orden de las preguntas y la logística de realización y recogida de la información de las entrevistas.

Tipos de entrevistas y aplicaciones especializadas. Las entrevistas cualitativas pueden adoptar varias formas: la entrevista informal tipo conversación, el acercamiento a través de una entrevista guiada y la entrevista estándar con un principio y un final. Todas estas formas varían según su grado de estructura, de planificación y la comparación de las respuestas en el análisis de datos. En la **entrevista informal** tipo conversación, las preguntas surgen a partir del contexto inmediato y se formulan según el curso natural de los hechos; no hay ninguna predeterminación en la formulación de las preguntas ni sobre su tema. Las conversaciones informales son una parte integrante de la observación participante. En la **entrevista guiada**, los temas se eligen antes; pero, el investigador decide el orden y la expresión de las preguntas durante la entrevista. Tanto la entrevista informal tipo conversación, como el acercamiento a través de una entrevista, se basan en la charla y en las circunstancias. En la **entrevista estándar con un principio y un final**, a los participantes se les plantean las mismas preguntas, en el mismo orden, con lo que se reduce la flexibilidad del entrevistador; sin embargo, la formulación estándar de las preguntas puede limitar la naturalidad y la congruencia de la respuesta (ver tabla 11.4).

La selección de la estrategia para entrevistar depende del contexto y del objetivo: (1) obtener las impresiones actuales de actividades, funciones, sentimientos,

TABLA 11.4: TIPOS DE ENTREVISTAS

Tipo de entrevista	Descripción
Entrevista informal, tipo conversación	Las preguntas surgen del contexto inmediato. No hay temas predeterminados ni nada escrito.
Entrevista guiada	Los temas se eligen antes. El investigador decide el orden y la formulación de las preguntas durante la entrevista. La encuesta de la entrevista puede aumentar la comprensión.
Entrevista estándar con principio y final	La formulación de las preguntas y el orden en que se presentan están predeterminados.

motivaciones, preocupaciones y pensamientos; (2) obtener expectativas para el futuro o experiencias anticipadas; (3) verificar y ampliar la información obtenida a partir de otras fuentes; y (4) verificar y ampliar las corazonadas e ideas desarrolladas por los participantes o el investigador.

Las aplicaciones especializadas de la estrategia de entrevista son entrevistas con información clave, entrevistas sobre la carrera profesional y sobre la vida, entrevistas de casos atípicos y entrevistas fenomenológicas. Las **entrevistas a informadores decisivos** son entrevistas en profundidad a individuos que tienen conocimientos especiales, situación o cualidades comunicativas que están deseando compartir con el investigador. Normalmente, se escogen porque tienen acceso a observaciones inasequibles para el etnógrafo. A menudo, se trata de individuos atípicos y deben seleccionarse cuidadosamente entre otros posibles informadores clave¹.

Las **entrevistas sobre la carrera profesional y la vida**, de las que se obtienen narraciones sobre la biografía de individuos, son empleadas por los antropólogos para obtener datos sobre la cultura. Los etnógrafos relacionados con la educación utilizan esta técnica para entrevistar sobre la vida o la carrera profesional. Por ejemplo, cuando un examen sobre la noción de carrera profesional de profesoras de primaria difiere de la investigación principal sobre la carrera profesional de los hombres, el investigador sugiere que el concepto de carrera profesional debe englobar a las mujeres trabajadoras (Biklen, 1985). La investigación sobre la carrera profesional y la vida de los educadores requiere frecuentemente, de dos a siete horas de entrevistas y puede llevar un tiempo considerable reunir a los informadores si la experiencia social compartida sucedió hace años.

Las **entrevistas de casos atípicos** son una aplicación especial de la entrevista que se centra en personas consideradas influyentes, destacadas y bien informadas dentro de una determinada organización o comunidad. Normalmente, los casos atípicos están familiarizados con la visión global de la organización, las relaciones que ésta establece con otras organizaciones y, especialmente, la estructura legal y financiera de la organización. El investigador debe confiar en el patrocinio, las recomendaciones y las presentaciones para conseguir citas con los casos atípicos. Frecuentemente, los casos atípicos prefieren una interacción más activa con el entrevistador, por lo que el grado de control de la entrevista variará mucho más. Los casos atípicos responden bien a interrogatorios sobre áreas de contenido amplio y a preguntas provocadoras e inteligentes que les permiten usar sus conocimientos libremente. A menudo, los casos atípicos aportan sentimientos e intenciones porque se sienten cómodos en el ámbito de las ideas, los principios y las generalizaciones².

¹ Ver J. P. Spradley (1979), *The Ethnographic Interview* para un debate detallado sobre las entrevistas a personas clave.

² Ver L. Dexter (1970), *Elite and specialized interviewing* y R. Gorden (1981), *Interviewing: strategies, techniques, and tactics*.

Las **entrevistas fenomenológicas** son un tipo específico de entrevista en profundidad utilizadas para estudiar el significado o la esencia de una experiencia vivida entre unos participantes seleccionados. La estrategia puede ser una única y larga entrevista completa con cada persona o tres entrevistas separadas con cada una de ellas³. Los estudios fenomenológicos investigan lo que se experimenta, cómo se experimenta y, finalmente, los significados que los entrevistados dan a la experiencia. La experiencia que se estudia es, normalmente, algo que ha afectado profundamente al individuo, como por ejemplo recordar un embarazo adolescente, incestos durante la niñez o adquirir una deficiencia física. Antes de entrevistar, el investigador escribe una descripción completa sobre cada una de sus experiencias. Las entrevistas fenomenológicas permiten al investigador tener un núcleo explícito de su experiencia personal combinado con las experiencias de los entrevistados. Frecuentemente, los educadores hacen uso de las entrevistas fenomenológicas de manera general para obtener los múltiples sentidos de una experiencia.

Preguntas cualitativas, tanteos y pausas. El contenido de las preguntas varía debido a los diferentes objetivos y a los problemas de la investigación, las estructuras teóricas y la selección de los participantes. Plantear preguntas de la investigación principal, probablemente, no proporcione datos válidos para la entrevista; sin embargo, el análisis de diferentes alternativas es esencial en la elaboración escrita de la entrevista⁴. Las preguntas de la entrevista pueden centrarse en la experiencia o en el comportamiento, en opiniones y en valores, en sentimientos, en conocimientos, en sensaciones y en el pasado de los individuos o en información demográfica (ver tabla 11.5). Cada una de estas preguntas puede formularse en presente, en pasado o en futuro.

La realización de entrevistas cualitativas requiere formular preguntas de final abierto. Los investigadores principiantes empiezan, a menudo, por los datos que quieren obtener y formulan las preguntas de manera que los entrevistados puedan deducir la respuesta que se espera. Esto es lo que se llama *preguntas con respuestas previstas*, que consigue escuetos sí o no o frases cortas como respuesta. Cuando ocurre esto, la entrevista adopta un tono de interrogatorio más que de conversación.

Las entrevistas cualitativas en profundidad destacan más por sus tanteos y pausas que por los formatos especiales de las preguntas. La confianza, la espontaneidad, el mantener el contacto visual e indagar a través de la elección de las palabras y el tono de voz que el investigador «oye» y que lo conecta con la persona, proporciona datos más válidos que un acercamiento rígido. En el extracto 11.7, a partir de la primera pregunta y hasta la tercera, se recoge información y se establece relación con las madres de los barrios céntricos de la ciudad en un

³ Ver Moustakas (1994) y Seidman (1998).

⁴ Ver Denzin (1978); Lofland & Lofland, (1984); Patton (1990); Pelto & Pelto (1978); Schatzman y Strauss (1973); Spradley (1979).

TABLA 11.5: TIPOS DE PREGUNTAS DE ENTREVISTA

Tipo	Descripción e ilustración
Experiencia/ comportamiento	Lograr saber lo que alguien hace o ha hecho: descripciones de experiencias, comportamientos, acciones, actividades durante la ausencia del etnógrafo: «si hubiera estado aquí ese día, ¿qué experiencias le habría visto tener?».
Opiniones/ valores	Lograr saber lo que alguien piensa sobre sus experiencias, qué pueden revelar las intenciones de alguien, metas y valores: «¿qué te gustaría que ocurriera o qué crees acerca de...?».
Sentimientos	Lograr saber cómo reaccionan las personas emocionalmente ante sus experiencias: «¿te sientes preocupado, feliz, con miedo, intimidado, confiado en...?».
Conocimiento	Conseguir información demostrable que la persona tiene o que considera la persona como demostrable.
Sensaciones	Lograr conocer las descripciones de la persona sobre qué y cómo ve, oye, toca, saborea y huele en el mundo que le rodea: «¿qué te preguntó el consejero cuando fuiste a su oficina? ¿Cómo te recibió?».
Pasado / información demográfica	Lograr conocer la descripción que la persona tiene de ella misma para ayudar al investigador en la identificación y localización de las personas en relación a otra gente: «Información habitual, educación, ocupación, residencia / movilidad y cosas por el estilo».

Pueden formularse en presente, en pasado o en futuro. Ver Patton (1990).

programa de intervención con la infancia. Las preguntas 4, 5 y 6, frecuentemente, se convertían en un solo tema debido a algunas madres; para otras, la elección de las palabras en cada pregunta les condujo a aportar información diferente. Después de una serie de entrevistas, los investigadores normalmente encuentran más fácil ajustar la entrevista a cada persona.

EXTRACTO 11.7

PROTOCOLO DE LA ENTREVISTA CON PREGUNTAS ALTERNATIVAS Y TANTEOS

1. Dígame cómo se enteró de que iba a celebrarse el programa infantil _____.

Preguntas alternativas y tanteos

¿Quién le habló del programa infantil?

¿Cómo entró a formar parte del programa?

¿Cómo se enteró de la celebración del programa?

...

(continúa)

(continuación)

4. ¿Qué tipo de servicios recibe su hijo/a y su familia del programa infantil? ¿Cómo se tomó la decisión de que su hijo/a y su familia recibirían dichos servicios?

Preguntas alternativas y tanteos

- ¿Qué hizo el programa infantil por usted (con usted)?
- ¿Qué hizo el programa infantil por su hijo/a?
- ¿Qué hizo el programa infantil por su familia?
- ¿Quién decidió lo que el programa infantil haría por ustedes?
- ¿Tiene usted algo que ver en la decisión de lo que el programa infantil haría?
- ¿Otros miembros de su familia estuvieron relacionados con dicha decisión?
- ¿Sintió que sus peticiones fueran escuchadas?

5. ¿Qué tipo de control o energía sintió (siente) haber recibido a partir de los servicios del programa infantil?

Preguntas alternativas y tanteos

- ¿Quién tomó la decisión sobre lo que el programa infantil haría por usted?
- ¿Sentía que podía tomar decisiones sobre los servicios que el programa infantil le proporcionaba?
- ¿Hubo diferencias entre lo que usted esperaba que fuera el programa y lo que fue realmente? ¿Cuáles?
- Cuando el personal trabajaba con el /la niño/a, ¿era suyo el niño o del programa infantil?

6. ¿Qué resultó ser de gran ayuda de los servicios que su hijo/a y familia recibieron del programa infantil?

Extensiones / tanteos

- ¿Cómo consiguió que el programa realizase estas cosas?
- ¿Quién lo hizo?
- ¿Quién decidió hacerlo?

Fuente: Gamel-McCormick, M. T. (1991), *Talking, teaching, and telling*, informe de evaluación no publicado. Virginia Commonwealth University: School of Education, Richmond. Apéndice B.

Las técnicas para obtener unas buenas preguntas cualitativas son las críticas sobre entrevistas escritas por entrevistadores experimentados, una guía sobre entrevistas, y una revisión de las preguntas iniciales para su redacción final. A continuación encontramos ejemplos de formulación de preguntas con las repuestas de campo y la formulación final de una guía para la entrevista (Schumacher, Esham y Bauer, 1985, pp. 150-153).

Preguntas con respuestas iniciales previstas	Últimas preguntas cualitativas
P: ¿Tenían dificultades los profesores en los seminarios? R: Sí.	P: ¿Esperaba que los profesores tuvieran dificultades en el seminario?
P: ¿Los profesores cambiaron? R: Algunos.	P: ¿Cómo afectó a los profesores su participación en el seminario?
P: ¿Aprendió usted algo enseñando en los seminarios? R: Sí.	P: ¿Qué aprendió sobre las estrategias de enseñanza que utilizó con el grupo?
P: ¿Identificó algún problema del que el Comité de Planificación debería ocuparse? R: Sí	P: ¿Qué le gustaría que hiciera el Comité de Planificación?

Aunque lo que acabamos de leer sea un ejemplo extremo de respuestas a preguntas con respuestas previstas, resulta evidente que las preguntas cualitativas generarían (y, de hecho, lo hacen) datos diferentes, obteniendo múltiples datos sobre los seminarios.

Las preguntas con respuestas previstas pueden ser *preguntas guiadas*, lo que implica una respuesta predeterminada. Guiar preguntas puede expresarse como pregunta predeterminada, una pregunta que implica una presunción deliberada para provocar una respuesta compleja o elaborada. Los ejemplos en la columna de la izquierda son preguntas con respuestas previstas que han sido reformuladas como preguntas con respuestas presupuestas en la columna de la derecha. (Schumacher, 1984b, p. 75-82).

Preguntas con respuestas previstas	Preguntas con respuestas presupuestas
¿No les molestó a los profesores que estaban trabajando, recibir clases después de las horas de escuela?	¿Cuáles fueron los aspectos más difíciles [del programa] de llevar a cabo? (Presuposición: hubo muchas dificultades)
¿Esperaba que los profesores cambiaran tras haber participado en los seminarios?	¿Qué cambios esperaba que los profesores experimentaran a partir de su participación en los seminarios? (Presuposición: hubo un cambio inmediato)
¿Sabe si se ha producido algún resultado o efecto inesperado?	¿Cómo manejó el Comité de Planificación las oportunidades inesperadas? (Presuposición: hubo oportunidades inesperadas)

Algunos investigadores subrayan la ineficacia de las preguntas introducidas por la partícula interrogativa *por qué*. Normalmente, estas preguntas parten de conjeturas, frecuentemente son ambiguas y demasiado abstractas para obtener datos concretos. No obstante, en algunas situaciones, empezar con la partícula interrogativa *por qué* permite al investigador obtener procesos causa-efecto o relaciones que son potencialmente informativas ya que revelan lo que la persona da por sentado.

La secuencia de preguntas. El que una entrevista sea efectiva o no, depende de la eficiencia de las indagaciones y el orden de las preguntas. Sugerimos algunas líneas directrices a continuación.

1. Gracias a las ***indagaciones de las entrevistas*** obtenemos más detalles, más explicaciones y respuestas más claras. Las entrevistas bien diseñadas se redactan de manera que se pueda identificar el orden y la redacción de las preguntas necesarias para adaptar los temas según las respuestas de las personas. Las preguntas más amplias se redactan, con frecuencia, más concretamente como indagaciones. Encontramos un ejemplo en el extracto 11.7, el protocolo de una entrevista para padres que han apuntado a sus hijos en un programa de servicios de intervención. Los investigadores deben hablar menos que los participantes; las pautas que necesitan, normalmente, pueden reducirse a unas cuantas palabras durante la entrevista.
2. El ***planteamiento de los objetivos del investigador*** se realiza, normalmente, al principio. En este momento se expone la seguridad de la protección de la identidad de las personas y un vistazo general sobre los posibles temas de discusión. La información comunicada se refiere a la importancia de los datos, las razones de esta importancia y el consentimiento del entrevistador a explicar el objetivo de la entrevista al entrevistado. Los investigadores proporcionan explicaciones o recursos relacionados con el núcleo de la entrevista para que los participantes se adapten a las nuevas áreas que se irán tratando a lo largo de la entrevista.
3. El ***orden de las preguntas*** varía, aunque la mayoría de los investigadores elige un orden que les permita obtener de manera eficiente, los datos adecuados a cada pregunta. Si el orden de las preguntas es rígido, será más fácil comprender el funcionamiento de la entrevista, pero también será más cansado y aburrido para el participante y el entrevistador. Las preguntas de carácter general se agrupan por temas, pero en muchos casos, los entrevistadores ignoran el orden que seguirán las preguntas ya que depende de las réplicas de las personas y el rumbo que siga la entrevista.
4. Las ***preguntas demográficas*** pueden aparecer a lo largo de toda la entrevista o presentarse en las conclusiones finales. Algunos investigadores prefieren obtener estos datos al principio de la entrevista para establecer una buena relación con los participantes.

5. Las *preguntas complejas, delicadas o difíciles* se reservan normalmente para la mitad o el final de la entrevista cuando el interés del participante ha aumentado. Algunos entrevistadores prefieren empezar las entrevistas con preguntas descriptivas y orientadoras y, poco a poco, van abordando temas más complejos sobre opiniones y explicaciones.

Logística de las entrevistas. Los etnógrafos escogen temas para la entrevista y preguntas mientras se planifica la logística general que influye en una sesión de entrevista. Cinco imprevistos que pueden afectar la sesión de una entrevista son (1) la *duración* de la sesión; (2) el *número*, o cuántas entrevistas separadas se necesitan para obtener los datos; (3) el *lugar* de la entrevista; (4) la *identidad de los individuos* que participan en el estudio y el número que estará presente en la sesión, y (5) las *formas de informar* o cualidades comunicativas de los entrevistados. Algunos diseños de investigación se conciben para entrevistas periódicas y otros diseños requieren entrevistar sólo después de acontecimientos importantes.

Los entrevistadores varían sus estilos de intervenir. Los modos interactivos pueden ser de confrontación, neutrales desde un punto de vista emocional, pero sofisticados o empáticos desde un punto de vista cognitivo. Pueden utilizarse técnicas específicas para tranquilizar, mantener el control de la entrevista y usar apropiadamente el apoyo y el reconocimiento. Los entrevistadores más cualitativos prefieren un tono coloquial para demostrar empatía y comprensión al tiempo que transmiten aceptación para así poder conseguir datos sutiles y válidos.

Las grabaciones de las entrevistas, transcripciones y elaboraciones. Los datos primarios de las entrevistas cualitativas son relatos recogidos, palabra por palabra, sobre lo que ocurre en la sesión de la entrevista. La grabación en cinta de la entrevista garantiza la integridad de la intervención verbal y proporciona material para la fiabilidad en los controles. Estas ventajas pueden ir en detrimento de la confianza del participante y suponer un fracaso automático. El uso de la grabadora *no* implica la desaparición de la necesidad de tomar notas para reformular preguntas y para anotar comunicación no verbal, aspecto que facilita el análisis de los datos. En muchas situaciones, las notas tomadas por el mismo entrevistador pueden ser el mejor método. El entrevistador que graba se ve obligado a estar atento, la grabación puede ayudar a la marcha de la entrevista y garantiza la autenticidad de lo que se ha escrito (al principio del análisis de datos) durante la entrevista. Sin embargo, ni la toma de notas ni la grabación en cinta deben entorpecer la atención del investigador respecto a la persona.

Inmediatamente después de la entrevista, el investigador completa y escribe en ordenador las notas o la cinta. Los borradores tendrán que editarse para corregir los errores de mecanografía o transcripción y, una vez hecho esto, se organizarán según se quiera para obtener el informe final. El resultado final contiene los datos literales y las anotaciones de la comunicación no verbal con pistas y comentarios iniciales para empezar la investigación. El informe final también

incluye la fecha, el lugar y la identidad o código del informante. El extracto 11.8 ilustra la obtención de datos a partir de entrevistas en profundidad.

EXTRACTO 11.8

TRANSCRIPCIONES DE LAS ENTREVISTAS – CITAS SELECCIONADAS

[Experiencia de los directores con dirección compartida en sus escuelas]

- *Siempre te preguntas si realmente tu papel es necesario. Piensas, bueno, ¿realmente me necesitan aquí, o quieren que esté aquí? Muchas veces empiezas a cuestionarte si este lugar podría seguir su curso sin ti. Entonces, llama un padre que no quiere volver a hablar con un profesor y en ese momento sé que se me necesita.*
- *Creo que los profesores no se dan cuenta de lo mucho que me necesitan; a veces piensan que si yo no estuviera aquí, ellos no tendrían tantas limitaciones. Ahora desempeño un papel más amplio que el de relaciones públicas y paso más tiempo comunicándome con el entorno...*
- *Estoy madurando. Aprendiendo a convertirme más en un compañero, aprendiendo a volverme más democrático. Desde luego, todavía tengo mucho que aprender. Se tiene mucho más trabajo en grupo que si intentas trabajar por tu cuenta...*
- *La escuela está bajo un mejor control precisamente porque no lo estoy controlando todo. Hablamos de estar... todos juntos en esto. Hay más intención de compartir, no menos...*
- *Personal y profesionalmente, [la dirección compartida] me ha proporcionado un fantástico sentimiento de satisfacción. Me gusta ver crecer a la gente. Me gusta estar en la frontera, en el territorio inexplorado. Cuando compartes el gobierno, avanzas en temas educativos y del currículo.*

Fuente: Blase, J y Blase, J (1999), «Shared Governance principals: The inner experience», p. 83-85, *NASSP Bulletin*, 83(606), 81-90.

Los investigadores redactan **elaboraciones de entrevistas** de cada sesión, reflexiones propias sobre su cometido y su relación, sobre las reacciones de los entrevistados, sobre información adicional y ampliaciones de los significados de la entrevista. Esta actividad es un momento importantísimo para llevar a cabo un control de calidad de los datos válidos. Muchas ideas iniciales, ya desarrolladas, serán revisadas más tarde junto con otras actividades de recolección de datos. Por regla general, por cada hora de entrevista, el investigador dedica cuatro horas de trabajo adicional para obtener las anotaciones finales y otras elaboraciones.

RECOGIDA DE DOCUMENTOS Y ARTEFACTOS

La selección de artefactos es una estrategia no interactiva para obtener datos cualitativos con una reciprocidad pequeña o inexistente entre investigador y participante. La selección de artefactos es menos reactiva que las estrategias

interactivas de las que el investigador no extrae conclusiones. Durante la estancia en el campo en escuelas, por ejemplo, los etnógrafos deben actuar recíprocamente con individuos –incluso aunque sólo sea de manera no verbal– y convertirse, en cierto modo, en participantes. Esto no es un impedimento, si el investigador anota las consecuencias de este papel interactivo. Por el contrario, las estrategias de selección de artefactos son no interactivas pero pueden requerir trabajo imaginativo en el campo para poder conseguir datos relevantes.

Las selecciones de artefactos son manifestaciones tangibles que describen las experiencias de la gente, su conocimiento, sus acciones y sus valores. Los investigadores cualitativos que estudian grupos contemporáneos, han adoptado las técnicas de los historiadores que analizan documentos (ver el capítulo 13) y de los arqueólogos que examinan objetos elaborados por pueblos antiguos.

Tipos de artefactos

Los artefactos de los grupos actuales y de las instituciones educativas pueden adoptar cuatro formas: documentos personales, documentos oficiales, objetos y medidas de desgaste (ver tabla 11.6)

Documentos personales. Los documentos personales son cualquier narración en primera persona que describe las acciones, las experiencias y las opiniones del individuo. Los documentos personales incluyen diarios, cartas personales y anotaciones anecdóticas. Normalmente, el investigador descubre estos documentos, pero a veces el etnógrafo pedirá a un participante que las registre en forma de anotaciones, diario, notas de preparaciones de lecciones o los apuntes de los padres sobre el desarrollo de un niño. Dichos documentos pueden aparecer durante una entrevista o gracias a la observación participante.

Documentos oficiales. Los documentos oficiales abundan en cualquier organización y pueden adoptar muchas formas distintas. Los informes, papeles de trabajo y borradores de propuestas, son documentos informales que proporcionan una perspectiva interna de la organización. Estos documentos describen las funciones y valores y cómo varias personas definen la organización. Los documentos internos pueden mostrar la cadena oficial de mando y proporcionar pistas sobre el estilo de la dirección y su trascendencia. Los documentos empleados para la *comunicación externa* son los producidos para el consumo público, informes, folletos de programas, partes sobre los horarios escolares, textos públicos y nuevas publicaciones. Estos documentos sugieren la perspectiva oficial de un tema o proceso. Los horarios escolares desde 1915 hasta 1980, por ejemplo, constituyeron una importante fuente para el análisis de una escuela innovadora tras un estudio etnográfico, y las solicitudes de trabajo del personal de la escuela proporcionaron pistas demográficas para la localización de estas personas para las entrevistas.

TABLA 11.6: SELECCIONES DE DOCUMENTOS Y ARTEFACTOS

Tipo	Ejemplos	Usado para
Documentos personales	Diarios Cartas personales Anotaciones anecdóticas	Perspectiva personal
Documentos oficiales	Papeles internos Comunicación externa. Anotaciones de estudiantes y expedientes del personal. Datos estadísticos (enumeración).	Perspectiva informal u oficial dentro de la organización. Perspectiva oficial para el público. Perspectiva de niños y empleados sobre la institución. Sugiere direcciones, plantea preguntas, corrobora descubrimientos cualitativos, describe protocolos y valores.
Objetos	Símbolos Objetos	Sugiere significados sociales y valores. Sugiere significados sociales y valores.

Los archivos existentes y las colecciones demográficas pueden ser localizados durante la permanencia en el campo y, normalmente, están disponibles para el investigador. Las instituciones también guardan anotaciones individuales de cada estudiante y empleado; normalmente, para lograr el acceso a dichos documentos es necesario el permiso de los padres, del estudiante o del empleado. Los expedientes de los alumnos y del personal pueden llegar a ser muy elaborados y pueden contener anotaciones e informes variados. Un expediente de alumno puede contener anotaciones sobre exámenes, asistencia a clase o comentarios anecdóticos de profesores, información de otras delegaciones y perfil familiar. Los investigadores usan los expedientes, no tanto, para conocer lo que dice sobre el estudiante sino más bien para averiguar lo que el expediente sugiere sobre las personas que realizaron las anotaciones. El expediente representa perspectivas diferentes (de psicólogos, de profesores, de asesores, de directores) sobre el estudiante.

Los *datos estadísticos* pueden constituir información demográfica sobre un grupo de población, índice de abandono de estudios, resultados alcanzados, número de actos violentos y despidos, anotaciones sobre asistencia a clase, número de estudiantes con deficiencias, listas de elecciones a los programas federales, número de accidentes deportivos y otros datos numéricos. Los investigadores cualitativos emplean los datos estadísticos de varias maneras: (1) para sugerir direcciones, (2) para proponer nuevas preguntas o (3) para corroborar datos cualitativos. Los investigadores cualitativos se interesan más por lo que dicen las estadísticas acerca de las suposiciones de las personas que las usan y recopilan

los datos: cómo las estadísticas revelan el pensamiento de la gente y el sentido común. Rutinariamente, los datos numéricos expresados describen protocolos y valores sociales de una organización. Los trabajadores de campo rara vez toman los datos estadísticos por su valor, sino que se cuestionan sobre el proceso social que produjo los datos y cómo se emplearon.

Los objetos. Los objetos son símbolos creados y formas tangibles que revelan procesos sociales, significados y valores. Ejemplos de símbolos son logotipos y mascotas de escuelas, de equipos y de clubes; ejemplos de objetos son tablas de gimnasia y trofeos, diplomas y pósters. En un estudio de colaboración institucional, una anotación simbólica fue el nuevo logotipo que combinaba parte de los emblemas de la universidad y del sistema escolar, para representar una nueva relación entre las dos organizaciones. Los datos interactivos revelan las dificultades superadas para obtener la aprobación oficial del uso de emblemas institucionales de un modo nuevo. Los datos obtenidos al año siguiente describen el uso del nuevo logotipo. Los investigadores cualitativos pueden investigar el valor del trabajo de profesores y estudiantes revisando periódicamente la media de las notas en las clases de primaria y corroborando estos datos con otros datos de campo.

Análisis e interpretación de la selección de artefactos

La selección y el análisis de artefactos requieren cinco estrategias:

1. La **localización de los artefactos** empieza con la entrada en el campo y continúa a lo largo del estudio. Los investigadores cuentan anticipadamente con artefactos y proceden a la localización y obtención de documentos y objetos. Los participantes también ofrecen documentos y artefactos.
2. La **identificación de los artefactos** requiere situar el artefacto de manera que se pueda acceder de nuevo a él y que sea recuperable. Los documentos se fotocopian; los objetos se fotografían o se graban. Las identificaciones se realizan anotando la categoría del artefacto, una breve descripción del artefacto, una historia de su uso y propietarios/sucesores y datos de aparición y representatividad.
3. El **análisis de los artefactos** requiere datos descriptivos sobre la producción o adquisición del artefacto. Preguntas importantes son quién lo usa, cómo lo usa, dónde lo usa y con qué propósito.
4. La **crítica de los artefactos** es la determinación de su autenticidad y exactitud para identificar los significados del artefacto en un escenario social.
5. La **interpretación del artefacto** debe ser corroborada por la observación y por los datos de la entrevista. La interpretación del artefacto para significados sutiles depende del contexto social y de otros datos.

OBSERVACIONES DE CAMPO Y TÉCNICAS SUPLEMENTARIAS

Una técnica fundamental de toda la investigación cualitativa es la observación de campo: consideraciones directas de la acción social de todos los días y lugares que se convierten en notas de campo. Una amplia variedad de técnicas suplementarias se utiliza, también, en la mayor parte de los estudios. Las técnicas suplementarias son las seleccionadas para ayudar a interpretar, elaborar y corroborar los datos obtenidos a partir de la observación participante, de las entrevistas en profundidad, de los documentos y de los artefactos.

Observación de campo

Las observaciones de campo cualitativas son anotaciones descriptivas en forma de notas de campo sobre los acontecimientos, personas, acciones y objetos en lugares. La observación de campo es una parte integrante de la observación participante y de las entrevistas en profundidad. Al principio, el investigador se conforma con una observación cuidadosa a medida que va explorando varias áreas de interés del lugar, buscando pautas de comportamiento y relaciones. En algunos escenarios, el observador participante recopila todos los datos mediante la observación de la escena en acción, como una clase o una reunión. El entrevistador, también, realiza apuntes de campo sobre el lenguaje corporal no verbal del entrevistado y sus gestos, que ayudan a interpretar los datos verbales. Las anotaciones de campo incluyen descripciones del contexto de la entrevista, en una oficina, un restaurante, una casa o cualquier otro lugar.

Técnicas suplementarias

Las técnicas suplementarias incluyen técnicas visuales, análisis de la comunicación no verbal, medidas de desgaste, estudios especiales y la observación de los grupos. Se trata de métodos independientes y especializados con estudios metodológicos publicados. No obstante, el investigador cualitativo los utiliza para corroborar los descubrimientos iniciales y para plantearse nuevas preguntas.

Técnicas visuales. Las técnicas visuales incluyen películas y fotografías de una escena social habitual. Las películas son particularmente útiles para confirmaciones. Las películas nos documentan sobre comportamientos no verbales y comunicación y pueden constituir un dato permanente. Las películas, sin embargo, pueden presentar problemas a la hora de interpretarlas. Se debe considerar la intrusión técnica, la selección de escenas y los gastos.

Análisis de la comunicación no verbal. El análisis de comunicación no verbal es muy importante en la mayoría de los estudios cualitativos. El estudio del movimiento corporal y de los mensajes que se desprenden de éste se llama *quinesiol-*

gía. Las anotaciones sobre expresiones faciales, gestos y movimientos pueden adoptar la forma de datos verbales. Un entrevistador confía más en las respuestas si el lenguaje corporal es congruente con lo que se dice. Sin embargo, muchos gestos tienen significados distintos en diferentes culturas. El estudio sobre el uso que la gente hace del espacio y su relación con la cultura se llama *proxemia*⁵. Los estudios se han llevado a cabo tomando como base el uso del espacio interpersonal en lugares públicos y la identificación de los hábitos territoriales de ciertas culturas. Los investigadores cualitativos pueden anotar cómo reaccionan los demás con respecto al espacio y ante la invasión de la privacidad (territorio personal) en los asientos asignados de clase, de las áreas de trabajo con mucha gente y la selección de asientos en reuniones formales. Debemos tener cuidado con las interpretaciones de la comunicación no verbal porque tales datos deben entenderse junto con otros datos.

Medidas de desgaste. Las medidas de desgaste indican el grado de uso selectivo, por parte del grupo, de algunos materiales. Ejemplos de medidas de desgaste son las trayectorias que los estudiantes siguen una vez que salen de la escuela o el uso de baldosas de vinilo en determinados vestíbulos como sugerencia de la frecuencia de uso y posibles reuniones sociales. Aunque el desgaste es la medida, la interpretación de las tasas de desgaste proviene de una comprobación con las anotaciones sobre mantenimiento de suelos y edificios, la plantación de césped y la colocación de tejas durante los últimos años.

Estudios especiales. Los instrumentos para llevar a cabo los estudios pueden tomar la forma de estudios de confirmación, de pruebas elaboradas por participantes e, incluso, de técnicas proyectivas con fotos, dibujos y juegos. Los datos recogidos durante los nueve meses de observación del participante, por ejemplo, pueden corroborarse con un cuestionario administrado a los directores y a los participantes elaboradores de pruebas durante los recesos planeados y los talleres (Schumacher y Esham, 1986).

La observación de grupos. Una variación de una entrevista es la observación de los grupos de entrevistas (FGI, *focus group interview*), una estrategia para obtener una mejor comprensión del problema o una valoración del problema, asunto, producto nuevo, programa o idea a través de la entrevista a un grupo de personas y no a través de cada persona individualmente. Con la creación de un marco social en el que a los miembros del grupo se les estimula mediante las percepciones e ideas del vecino, se incrementa la calidad y la riqueza de los datos gracias a una estrategia más eficaz que las entrevistas individuales⁶. Los observadores

⁵ Ver E. T. Hall (1959, 1966, 1974).

⁶ El FGI se utilizó, en primer lugar, como estudio de evaluación y de planes, sin embargo, los sociólogos desarrollaron su enfoque original (Lofland, 1984; Scahtzman y Strauss, 1973). El FGI se utiliza, a menudo, en investigación de mercados y presta servicios de cierta utilidad en el cuidado de la salud.

participantes y las entrevistas en profundidad utilizan la observación de grupos como una confirmación de las técnicas. La investigación de estudios de caso y los estudios críticos pueden utilizar la observación de los grupos como una técnica más.

Si añadimos técnicas suplementarias a un estudio, aumentamos la validez de los descubrimientos iniciales y la credibilidad del estudio en su totalidad. Sin embargo, los investigadores más cualitativos no utilizan forzosamente todos y cada uno de los métodos para la obtención de sus datos.

NORMAS DE ADECUACIÓN PARA ESTRATEGIAS CUALITATIVAS

Muchos estudios cualitativos se publican en forma de libros o de informes en lugar de cómo artículos periodísticos. Con la creciente aceptación de la metodología, cada vez más periódicos publican manuscritos cualitativos. Los estudios que se publican en los diarios aparecen sumamente sintetizados o, incluso, en ocasiones sólo se habla de uno de los muchos descubrimientos realizados para adaptarse así al formato del periódico. El típico artículo periodístico puede también reducir los procedimientos metodológicos que deberían explicitarse en el estudio completo. Las siguientes preguntas ayudarán al lector:

Entrada en el campo

1. ¿Proporcionaron los problemas previstos suficientes criterios de selección para el lugar o el perfil de los individuos a los que se debía entrevistar?
2. ¿El cometido de la investigación está claramente articulado y es apropiado para las preguntas de la investigación?
3. ¿Cómo afecta el papel del investigador en la recogida de datos? ¿Cómo dirige el investigador sus posibles influencias?

La observación participante

1. ¿Cómo son las múltiples estrategias empleadas en la recogida de datos? ¿Cuál fue el método de base y cuáles se emplearon para corroborar los datos?
2. ¿La duración de la recogida de datos en el lugar fue detallada y razonable?
3. ¿Se presentan como datos las anotaciones de campo descriptivas?

Entrevistas en profundidad

1. ¿Las estrategias de muestreo tienen el objetivo de escoger a las personas debidas para la entrevista?
2. ¿Las personas entrevistadas quedaron ocultas a causa del atributo o perfil desarrollado por el estudio antes de proceder a las entrevistas?

3. ¿El tipo de entrevista seleccionado era apropiado para el problema de investigación?
4. ¿Los datos presentados indican que las preguntas de la entrevista eran apropiadas?

Técnicas suplementarias

1. ¿Se utilizaron apropiadamente las técnicas suplementarias para el estudio y dieron como resultado datos válidos?

RESUMEN

Los siguientes puntos resumen las características más importantes de las estrategias cualitativas:

1. La recogida de datos es el empleo de estrategias multimétodo; pero, sólo un método básico, como la observación participante o las entrevistas en profundidad, es seleccionado para el estudio.
2. Los problemas previstos determinan la dirección inicial y las estructuras conceptuales; sirven como guía del trabajo de campo pero no establecen ningún límite, ya que esto se producirá según el lugar y las circunstancias.
3. Los problemas previstos son reformulados durante la recogida de datos.
4. La selección del lugar se efectuará según los criterios de los problemas previstos, la adecuación y la viabilidad.
5. El observador participante, primero, traza el mapa de campo para adquirir una visión del contexto en su totalidad y para asegurar las muestras, obteniendo de este modo una selección de participantes con abundante información y escenarios sociales.
6. La observación participante es una permanencia prolongada en el campo que obtiene y corrobora consideraciones importantes desde diferentes perspectivas anotadas en forma de notas de campo.
7. Las entrevistas en profundidad varían el formato, los tipos de preguntas formuladas, el orden de las preguntas y el planteamiento de la entrevista.
8. Las anotaciones de las entrevistas son notas de campo, grabaciones en cassette, transcripciones y elaboraciones de entrevistas.
9. Las selecciones de artefactos son documentos personales, oficiales y objetos y deben ser corroborados junto con otras evidencias.
10. Las técnicas suplementarias se utilizan para verificar los datos recogidos a través de la observación participante o de las entrevistas en profundidad; pueden ser técnicas visuales, anotaciones de comunicación no verbal, medidas de desgaste, estudios especializados y la observación de grupos.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La selección de estrategias cualitativas empieza con un análisis de...
 - a. problemas de investigación previstos y preguntas.
 - b. una hipótesis de investigación.
 - c. una lista de variables.
 - d. una hipótesis estadística.
2. Las conclusiones obtenidas a partir de los problemas previstos indican que el investigador...
 - a. seguirá dichas ideas preconcebidas rígidamente.
 - b. observará sólo estos problemas anticipados.
 - c. reformulará estos problemas durante la recogida de datos.
 - d. lanzará teorías para estudios de verificación.
3. La entrada en el campo es guiada por...
 - a. la probabilidad de muestreo.
 - b. un análisis de los problemas previstos.
 - c. la solicitud de amigos para entrevistarlos.
 - d. la necesidad de obtener una descripción holística del contexto social.
4. ¿De entre los siguientes cometidos o funciones del investigador, cuál es inapropiado?
 - a. Observador participante.
 - b. Entrevistador.
 - c. Papel doble de participante e investigador.
 - d. Observador total o participante total.
5. Las estrategias para la recogida de datos incluyen...
 - a. la observación participante.
 - b. las entrevistas en profundidad.
 - c. la selección de documentos y artefactos.
 - d. todo lo anterior y las técnicas suplementarias.
6. La observación participante requiere descripciones de...
 - a. quién está en el grupo o en la escena social.
 - b. cuándo y dónde tiene lugar la escena social.
 - c. cómo actúa el grupo.
 - d. todo lo anterior es correcto.
7. Las entrevistas en profundidad normalmente son...
 - a. entrevistas a modo de conversaciones informales.
 - b. entrevistas que sirven de guía para realizar entrevistas.
 - c. entrevistas estándar de respuestas abiertas.
 - d. entrevistas con respuestas cerradas.

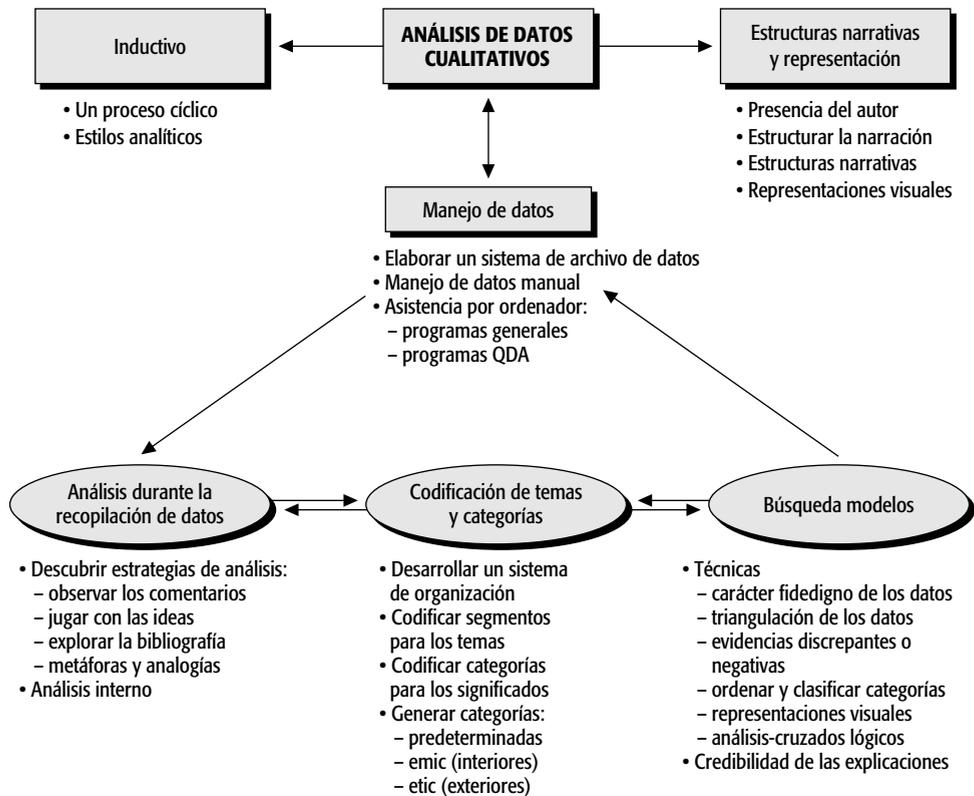
8. En las entrevistas en profundidad, los entrevistadores...
 - a. siempre graban en casete las sesiones.
 - b. rara vez llevan a cabo el protocolo de la entrevista.
 - c. evitan las preguntas alternativas y las indagaciones.
 - d. redactan las elaboraciones de entrevista para cada una.
9. Las entrevistas en profundidad...
 - a. son similares a interrogatorios.
 - b. animan a la participación del entrevistado a través de pausas y tanteos.
 - c. siguen un orden de preguntas inflexible.
 - d. plantean preguntas con respuesta previstas.
10. Los documentos y los artefactos...
 - a. son útiles para verificar otros datos.
 - b. se limitan a documentos personales.
 - c. son datos aceptados.
 - d. son disponibles raramente.

Problemas de aplicación

1. Un inspector solicita a un investigador que observe cómo una escuela de primaria lleva a cabo un nuevo plan de estudios científico para decidir si otras escuelas deberían adoptarlo. El investigador establece contacto fácilmente con el jefe del departamento de ciencias y con el director, y observa el plan de estudios en cinco de los seis profesores. El profesor número seis, que parece oponerse a tal innovación, sólo cuenta sus experiencias con el nuevo plan de estudios. Este profesor se las arregla para evitar enseñar el plan de estudios cuando el investigador está presente. ¿Qué debería hacer éste?
2. Un investigador vive en una residencia de estudiantes para averiguar cómo los estudiantes de instituto que acuden a un programa de escuela de verano, desarrollan su creatividad a través de la fotografía. Aunque, en un principio, el investigador pensó que la observación de las clases de fotografía y el programa de tarde bastarían, descubrió que las actividades sociales de los estudiantes durante el tiempo libre y las actividades extracurriculares durante el fin de semana influían en las producciones fotográficas de los estudiantes. ¿Debería observar y recopilar estos acontecimientos informales? ¿Serían útiles como fuentes de datos las fotografías de los estudiantes?
3. Durante la recogida de datos, un director de un programa de educación para adultos escucha las quejas presentadas por algunos adultos y ve que el investigador recoge tales datos. El director explica al investigador que tales observaciones, si llegaran a ser públicas, darían una imagen pobre del programa. El director solicita que se destruyan esas anotaciones particulares. ¿Cómo debería manejar la situación el investigador?

4. Esto es parte de un problema para un estudio propuesto: se han llevado a cabo muchos estudios sobre el impacto de las escuelas que optan por unir en las mismas clases a una minoría de estudiantes siguiendo medidas socio-psicológicas (autoestima, motivación, actitudes raciales, etc.) y según las medidas de los logros académicos (notas, resultados en los exámenes, nivel de conocimientos y de memoria y otros baremos). No obstante, la mayor parte de los estudios se centran en la valoración de las consecuencias cuantitativas de cada escuela con o sin la aplicación de este método y no en los procesos cualitativos del día a día escolar. Hay una ausencia de conocimientos sobre los procesos dentro del aula y del medio escolar y de la comunidad en la que estos procesos tienen lugar observando los patrones y dinámicas de socialización dentro de las escuelas agrupadas. ¿Parece el tema del estudio apropiado para un estudio cualitativo? Justifique su respuesta.
5. Reformule las siguientes preguntas guía de entrevista con preguntas cualitativas, y ordénelas adecuadamente para conseguir las impresiones de los profesores con motivo de la evaluación realizada por el director.
 - a. ¿Piensa que los profesores como profesionales deberían ser evaluados por sus directores?
 - b. ¿Ha visitado su director varias veces su clase antes de su evaluación anual?
 - c. ¿Mantén una conferencia usted y su director tras cada visita?
 - d. ¿Se corresponde la evaluación del director con su manera de enseñar?

Análisis de datos cualitativos



PALABRAS CLAVE

análisis inductivo	categoría
cristalización	categoría emic
metáfora	categoría etic
análisis interno	modelo
analogía	triangulación
codificación	representaciones visuales
comparación constante	manejo de datos
segmento	sistema de archivo de datos
tema	contexto
código	

Suponga que ha recogido cuidadosamente una serie de observaciones de campo sobre reuniones de planificación de profesores en una escuela, ha entrevistado a un número de niños sobre lo que ellos consideran un «buen» profesor o ha reunido las inquietudes del director sobre un tema polémico. ¿Cómo dar sentido a todos estos datos, que se despliegan a lo largo de numerosas páginas de anotaciones de campo y /o de transcripciones de entrevistas y documentos? ¿Por dónde empezar? ¿Cómo lo organizaría para colocar los descubrimientos importantes? ¿Qué técnicas utilizarían los investigadores? ¿Cómo ayudarían los programas de software en el tratamiento y análisis de datos?

El análisis de datos cualitativos es, en primer lugar, un proceso inductivo de organización de datos en categorías e identificación de modelos (relaciones) entre las categorías. La mayor parte de las categorías y de los modelos surgen a partir de los datos; más que aparecer impuestos por los datos recopilados, a partir de la formulación de datos. Los estilos analíticos entre investigadores van desde los estructurados hasta los surgidos a partir de intuiciones. No obstante, existen procesos generales y algunas técnicas que son comunes. La mayor parte de los investigadores emplean un estilo interpretativo/subjetivo antes que un estilo técnico/objetivo.

Este capítulo nos presenta el análisis cualitativo como un proceso de análisis intermedio y revelador, que desarrolla temas y categorías que pueden provenir inicialmente de los datos o ser predeterminados, y modelos de búsqueda para posibles explicaciones. Las técnicas de las que los investigadores hacen uso en cada fase del análisis de datos serán explicadas en este capítulo. Los datos se tratan a través de métodos manuales y por ordenador. Los programas del procesador de textos y los que están diseñados específicamente para el análisis de datos cualitativos, pueden ayudar en la organización de grupos de datos (notas de campo, transcripciones de

entrevistas y documentos) y en la agrupación de datos codificados. Los resultados se presentan con estilo narrativo. Las estructuras narrativas y las representaciones varían moderadamente según el modo de encuestar cualitativamente: etnográfico, fenomenológico, estudio de caso, teoría basada en hechos o estudios críticos.

ANÁLISIS INDUCTIVO: VISIÓN DEL CONJUNTO

El análisis de datos es un proceso cíclico que nunca se detiene y que forma parte de todas las fases de la investigación cualitativa (ver figura 12.1). El análisis inductivo significa que las categorías y los modelos surgen a partir de los datos más que venir impuestos por la formulación de datos. Los procesos inductivos dan lugar a una síntesis descriptiva de los datos más abstracta. El análisis de los datos no es la reducción de los mismos o un acercamiento cuantitativo, ni tampoco depende de los programas de software. Los programas de ordenador pueden ayudar en el proceso, pero no pueden reemplazar las actividades cognitivas del investigador.

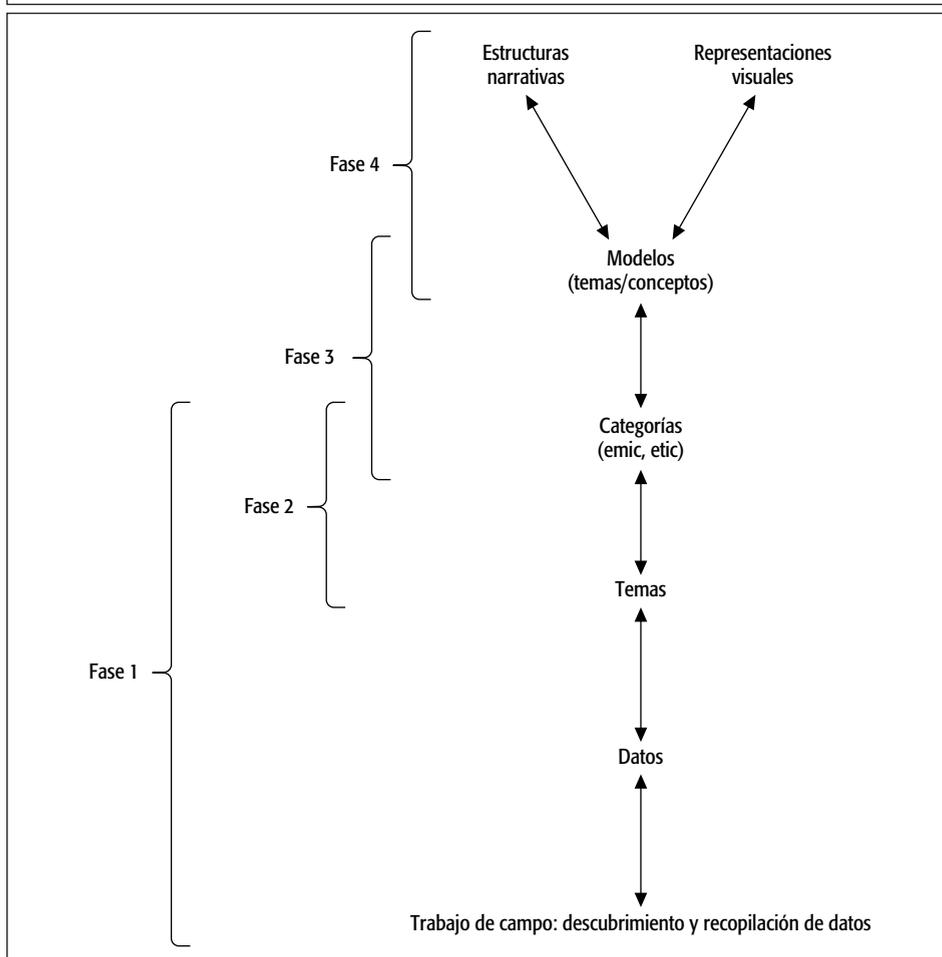
El análisis cualitativo es un proceso relativamente sistemático de selección, categorización, comparación, síntesis e interpretación, que nos proporciona explicaciones sobre el único fenómeno de interés. Los análisis de datos cualitativos varían mucho dependiendo de las diferentes investigaciones, objetivos, estrategias de formulación de datos y modos de encuestar cualitativos.

El proceso de análisis inductivo. El análisis de datos trae consigo varias fases cíclicas. El proceso general queda representado en la figura 12.1 en forma de cuatro fases que se superponen. No hay que pasar por alto que, a medida que el investigador avanza hacia niveles más abstractos del análisis de datos, tiene que volver constantemente al nivel básico de abstracción, y comprobar y pulir una y otra vez su análisis e interpretación. Y tampoco hay que olvidar que, a menos que algunos elementos estén presentes en los datos, el análisis no se realizará sin problemas. Los investigadores piden permiso para volver al campo, si resulta necesario, para recopilar más datos y dar por válidos los nuevos modelos que surgen.

La mayor parte de los investigadores cualitativos han aprendido que no existe ningún lugar en el que los procedimientos para la formulación de datos y para seguir la pista de estrategias analíticas sea estándar. Aunque el análisis de datos es inductivo, se utiliza un modo deductivo de pensar en momentos apropiados. «Dar sentido» a los datos depende totalmente del rigor intelectual del investigador y de la tolerancia para las tentativas de interpretación hasta que el análisis esté completamente finalizado.

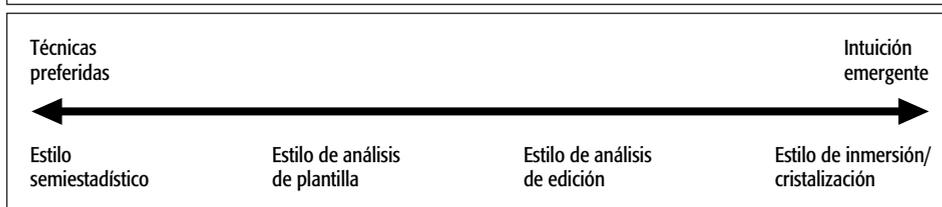
Los estilos analíticos. Los investigadores cualitativos desarrollan estilos analíticos; raramente explicitan todas las estrategias que utilizan para el análisis de datos. La figura 12.2 nos muestra una serie de estilos analíticos «idealizados» que van desde técnicas prefiguradas hasta intuitivas (Crabtree y Miller, 1999).

FIGURA 12.1: PROCESO DE ANÁLISIS DE DATOS INDUCTIVO



En el extremo de objetividad de la serie encontramos un estilo técnico y casi estadístico. El investigador decide las categorías por adelantado; las categorías quedan así predeterminadas y rígidas. Una estrategia de este estilo es contar las frecuencias de temas específicos o códigos. Un estilo plantilla, lógicamente, aplica los grupos derivados de códigos y las categorías a los datos, sin embargo, estas clasificaciones se revisan frecuentemente durante el análisis de datos. Los códigos iniciales provienen de preguntas de la investigación, temas en la guía de entrevista o categorías relevantes procedentes de la bibliografía. Este grupo inicial puede o no quedar retenido en el análisis final. El estilo de edición está menos prefigurado; el intérprete busca los grupos de datos por segmentos para

FIGURA 12.2: PROCESO DE ESTILOS ANALÍTICOS



Fuente: adaptado de Marshall y Rossman, 1999, p. 151.

ilustrar categorías de significado y escribir informes durante el proceso. Aunque el empleo de códigos es pequeño o inexistente, el analista debe agrupar los informes descriptivos que aclaren las mejores interpretaciones para establecer generalizaciones abstractas. En el estilo inmersión/cristalización, el investigador interrumpe la segmentación, la clasificación y la búsqueda de modelos durante un largo periodo de inmersión en los datos, sirviéndose de la intuición. La **crystalización** pretende ofrecer al analista la máxima cantidad de experiencias posibles dentro del estilo analítico. El investigador puede llevar a cabo varios análisis intensivos reflexivos simultáneamente. Este estilo implica, a menudo, volver a vivir cada experiencia de campo y cuestionarse constantemente sobre los datos con el fin de encontrar matices sutiles de significado. Los modelos se identifican a través de la reflexión iterativa.

La mayoría de los investigadores cualitativos tienden más a utilizar el estilo interpretativo/subjetivo y menos el técnico/objetivo. Analizar datos cualitativos es una actividad ecléctica (no existe una manera «correcta» y los datos pueden analizarse desde más de una manera). La mayoría de los investigadores cualitativos desean evitar los procesos estandarizados, porque el sello de la investigación cualitativa es el marco creativo del investigador. Cada analista debe encontrar su propio *estilo de artesanía intelectual*. Aunque no hay reglas estrictas que puedan seguirse, el investigador no se puede permitir poner límites a su inventiva. El análisis cualitativo debe llevarse a cabo ingeniosamente e, incluso, alegremente; pero también requiere una enorme cantidad de conocimiento metodológica y competencia intelectual.

El análisis empieza tan pronto como se recogen los primeros datos y va a la par que la formulación de datos porque cada actividad (formulación de datos y análisis interno) nos informa sobre otras actividades y nos conduce a ellas. Los datos se segmentan, es decir, se dividen en partes relevantes (unidades) o trozos de significado, escenas sociales o acontecimientos. Puesto que resulta difícil procesar grandes cantidades de contenido diverso al mismo tiempo, el analista se concentra en grupos de material más pequeños y parecidos. Sin embargo, el análisis intensivo empieza normalmente con la lectura de todos los datos para obtener

una visión del conjunto, lo que facilita la interpretación de las unidades de datos más pequeñas.

La principal herramienta intelectual es la comparación. La técnica de comparar y contrastar se usa en prácticamente todas las tareas intelectuales durante el análisis: identificar los segmentos de datos, dar nombre al tema / categoría y clasificarlos. Las categorías son aproximadas, al principio, y continúan siendo esquemas flexibles y no rígidos. El objetivo es identificar las similitudes.

Nuestra discusión ilustra una combinación del estilo de plantilla (empleo de códigos) con el estilo de edición, que muchos investigadores utilizan sin limitar los procesos de análisis subjetivo ni «obligar» a los datos a tomar la forma de predisposiciones. El estilo de plantilla se adapta fácilmente al análisis asistido por ordenador y puede dar como resultado una base experimental para el investigador y, de esta manera, desarrollar un estilo más interpretativo/subjetivo. Nosotros preferimos utilizar términos de análisis de datos cualitativos más generales en lugar de la terminología de las cinco tradiciones cualitativas del capítulo 2. Aunque el proceso de análisis de datos difiere en cierto modo y la terminología analítica varía debido a la tradición de la investigación¹, algunos principios generales y prácticas comunes guían a la mayor parte de los investigadores.

ANÁLISIS DEL DESCUBRIMIENTO EN EL CAMPO

El análisis del descubrimiento y el análisis intermedio tienen lugar durante la formulación de datos. La identificación y la síntesis de los modelos en los datos tienen lugar, normalmente, tras abandonar el campo. Presentamos estrategias a modo de *líneas directrices* para los análisis inductivos.

Análisis del descubrimiento

Puesto que los investigadores, a diferencia de los adivinos, no saben lo que se van a encontrar en el campo, seleccionan de forma secuencial las conclusiones finales de los estudios. Algunos planteamientos que facilitan la selección secuencial de la investigación son la elección de los problemas previstos; la elección de estructuras conceptuales; las estrategias de formulación de datos utilizadas en el estudio, incluyendo el muestreo, y el análisis intermedio.

¹ Por ejemplo, la fenomenología y la teoría argumentada poseen el mayor número de pasos explícitos en los análisis de datos. Lo que es aceptado como codificación abierta en la teoría argumentada es semejante a la primera etapa de la clasificación fenomenológica de *exposiciones o conjunto de categorías* en el caso de un enfoque fenomenológico del estudio.

Las estrategias del análisis del descubrimiento se emplean para desarrollar tentativas e ideas preparatorias durante la formulación de datos. Algunas de las estrategias que los investigadores emplean son las siguientes:

1. Escribir muchos «comentarios de observador» en las notas de campo y revisar los expedientes para identificar posibles temas, interpretaciones y preguntas. Los comentarios del investigador se separan de los datos reales y, generalmente, aparecen entre paréntesis. Una «regla de oro» es separar siempre los datos descriptivos de las interpretaciones que se desarrollan; la formulación de datos es una actividad descriptiva, mientras que los comentarios del investigador corresponden a una actividad reflexiva.
2. Escribir resúmenes de las observaciones y de las entrevistas para sintetizar y centrar en el tema. Frecuentemente, los investigadores se preguntan tras una visita de campo, «bueno, ¿qué acabo de aprender sobre mi tema? ¿Cómo puedo describir brevemente esto a alguien que no estaba presente? ¿Cuáles son los detalles importantes que se pueden contar a un compañero?». Al escribir resúmenes, el investigador obliga a su mente a seleccionar los aspectos importantes y a reorganizar dichos aspectos según un orden lógico. Puesto que los resúmenes de observaciones y de entrevistas constituyen un paso obtenido de los datos reales, estos informes se identifican claramente con el nombre de «resumen de...»
3. Jugar con las ideas, un proceso intuitivo, para desarrollar categorías iniciales sobre temas y conceptos. En ocasiones, puede adoptar la forma de diagramas exploratorios elaborados a partir de los datos que se van recogiendo. En otras ocasiones, adopta la forma de un intento de definición de la categoría.
4. Empezar a examinar la bibliografía y escribir cómo apoya o discrepa con las observaciones. Algunos investigadores sólo identifican posible bibliografía para leerla cuidadosamente más tarde porque la tarea más importante en esta fase es la formulación de los datos. Otros investigadores que recogen datos periódicamente examinan la bibliografía por lo que ya se ha escrito sobre el tema y por los conceptos útiles que les ayudan a comprender los datos.
5. Jugar con metáforas y analogías provisionales, no se trata de poner etiquetas, sino de hacer brotar ideas o capturar la esencia de lo que se observa y de las dinámicas de las circunstancias sociales. Una **metáfora** es una comparación sobreentendida entre cosas que esencialmente no se parecen entre sí para facilitar la comprensión de la situación. Los investigadores preguntan: «¿a qué me recuerda esto?». Por ejemplo, los niños con minusvalías en las escuelas podrían compararse con lo que se conoce como discriminación racial en las escuelas para descubrir las semejanzas y las diferencias. Una **analogía** es una metáfora que se hace explícita por medio de la palabra «como» para comparar dos fenómenos aparentemente distintos. Por ejemplo, en un estudio sobre la prevención del abandono de los

estudios, los papeles de los profesores se describieron con analogías de «poli de tráfico» o un cometido que mantiene a los estudiantes circulando dentro del sistema; y de «avestruz», que ignora la situación y espera que otro intervenga. Las metáforas y las analogías deben escogerse con cuidado, especialmente, para evitar posibles connotaciones racistas y sexistas. Sin embargo, utilizadas correctamente, las metáforas y las analogías iluminan los significados sutiles con una única frase.

Análisis intermedio

La mayor parte de los investigadores cualitativos realizan análisis intermedios regular y frecuentemente durante la formulación de datos. Los análisis intermedios cumplen dos objetivos: (a) para tomar decisiones durante la formulación de datos y (b) para identificar los temas que surgen y los modelos que se repiten. Los investigadores realizan análisis intermedios como si se tratara de una actividad constante de la formulación de datos, a menudo, tras cada tres o cinco visitas de campo o entrevistas, usando los grupos de datos recogidos (ver extracto 12.1).

A continuación comentamos algunas estrategias que utilizan los investigadores:

1. Examinar todos los datos recogidos hasta ese momento para localizar los posibles asuntos que contienen los datos. Aquí no debemos hacer hincapié en el significado del tema, sino en la adopción de una visión del conjunto sobre el alcance de los temas que obtenemos a partir de los datos.
2. Buscar los modelos que se repiten y que pueden convertirse en temas o modelos más significativos. Los temas se obtienen de las conversaciones y del lenguaje en las reuniones sociales, de las actividades que se repiten, de los sentimientos y los proverbios. También sugieren temas las observaciones del investigador extraídas de los comentarios del observador y de las elaboraciones de las entrevistas. Algunos modelos, es decir, algunas constantes y grupos de significado, resultan obvios en ese momento; otros se descubren tras haber realizado el análisis.
3. Volver a fijar los objetivos de la investigación para este análisis de datos y estudio particular. La mayoría de los datos cualitativos son tan extensos que se podrían elaborar a partir de ellos varios estudios. Una vez más, el investigador establece unos objetivos todavía más específicos y claros a través del análisis intensivo de los datos.

CODIFICACIÓN DE TEMAS Y CATEGORÍAS

Resulta prácticamente imposible interpretar datos a menos que los organice uno mismo. Los investigadores cualitativos combinan en un todo las operaciones

de organización, de análisis y de interpretación de datos, y llaman a ese proceso completo el «análisis de datos».

¿De dónde sacan las ideas los investigadores para organizar los datos? El investigador posee algunas ideas iniciales para la organización de los datos de cada experiencia de campo o para la planificación previa del estudio. Existen cinco fuentes que los investigadores utilizan para clasificar sistemas y organizar datos:

1. Las preguntas de la investigación y los problemas previstos o subpreguntas.
2. Los instrumentos de la investigación como una guía de entrevista.
3. Temas, conceptos o categorías manejados por otros investigadores en anteriores estudios.
4. Conocimientos previos del investigador.
5. Los datos en sí mismos.

EXTRACTO 12.1

ANÁLISIS INTERMEDIO EN LUGAR DE UNA CONCLUSIÓN

Puesto que todavía me encuentro en el proceso de análisis de los datos, me siento moralmente obligado a tratar los descubrimientos presentados aquí como lo que son: preliminares. Todavía hay muchas preguntas sin respuesta en mi investigación, por lo que sería pretencioso por mi parte presentar conclusiones en este momento. En lugar de ello, me gustaría que el lector considerara estos descubrimientos como indicadores sobre preguntas más amplias y mucho más complicadas que todavía no se han tomado en serio en el discurso escolar. Algunas de estas preguntas son:

¿Hasta qué punto, las jóvenes afroamericanas de esta localidad rural del sur, basan sus decisiones académicas en malas percepciones que se oponen a las valoraciones exactas de su situación educativa, social y económica? ¿Es posible establecer dicha diferenciación en primer lugar? ¿Qué dinámicas estaban en juego entre mis entrevistados y yo (como investigador) durante nuestra conversación sobre la creación de los valores educativos entre esta población de estudiantes y sus madres? Parece que, el largo y difícil camino hacia el mundo educativo para las mujeres estudiantes afroamericanas del mundo rural, acaba de empezar.

Fuente: Philipsen, M. (1993), «Values-spoken and values-lived: Female African Americans' educational experiences in rural North Carolina», *Journal of Negro Education*, 62 (4), 419-426.

No se nos tiene que pasar por alto que las cuatro primeras fuentes contienen categorías predeterminadas y que la quinta origina temas que se convierten en parte de una categoría más abstracta. Las cinco pueden usarse, pero el grado de eficacia de cada uno variará, dependiendo de la dirección y el objetivo del estudio.

La codificación es el proceso por el cual dividimos los datos en partes mediante un sistema de clasificación. Los investigadores desarrollan un sistema de clasificación al utilizar una de las tres estrategias siguientes:

- segmentación de los datos en unidades de contenido denominados temas (menos de 25-30) y agrupación de los temas en conjuntos más amplios para formar categorías;
- empezar con categorías predeterminadas de no más de cuatro a seis y dividir cada categoría en subcategorías más pequeñas;
- combinar las estrategias, utilizando algunas categorías predeterminadas y añadiendo nuevas categorías descubiertas.

Las estrategias son simplemente distintos puntos por los que empezar a crear un sistema de organización. Cualquier punto de partida empieza un proceso inductivo, generativo y constructivo porque *el grupo final de categorías no está totalmente predeterminado sino que está compuesto por datos de acuerdo con los significados de categorías*. Al dividir los datos en temas, los investigadores reorganizan los datos con el fin de trabajar con ellos. Normalmente, ninguna estrategia del principio aumenta la cantidad de material que es necesario procesar y, sólo más tarde, se agruparán las categorías para llevar a cabo la síntesis y la interpretación.

Esta técnica analítica se denomina comparación constante (Glaser y Strauss, 1967) –comparar y contrastar cada tema y categoría para determinar las características distintivas de cada uno–. Los investigadores se preguntan: «¿es éste el mismo tema que éste otro o son diferentes?». «¿En qué se diferencia o coincide este tema con los demás?». Los investigadores desarrollan categorías a partir de sus datos comparando constantemente cada categoría con otras para identificar sus particularidades distintivas. Se emplea una técnica similar en la búsqueda de modelos entre las categorías. Primero, describiremos el proceso para desarrollar un sistema de organización a partir de los datos, es decir, recurriendo a la estrategia número cinco.

Desarrollar un sistema de organización a partir de los datos

La mayor parte de los investigadores desarrollan un sistema de clasificación mientras recogen datos. Algunos investigadores empiezan cuando tienen una quinta o una cuarta parte de los datos, o al menos, tres grupos. Otros investigadores empiezan por el punto conveniente después de haber obtenido una visión general del fenómeno al completo.

El desarrollo de un sistema de organización consiste en dividir los datos en segmentos, es decir, en partes más pequeñas de datos que contengan una «porción del significado». Los fragmentos de los datos se denominan segmentos,

incidentes, unidades de significado o unidades de análisis. Un segmento de datos es comprensible por sí mismo y contiene una idea, episodio o asunto relevante para el estudio. Los segmentos pueden ser de cualquier tamaño: una palabra, una frase, unas cuantas líneas o varias páginas que contienen un acontecimiento entero o las explicaciones de los participantes con varias unidades más cortas dentro del segmento.

Cada segmento posee dos contextos. El primero es el conjunto de datos en el que se sitúa el segmento: el campo particular de observación o entrevista. El segundo contexto es la «reserva de significados» (Marton, 1986) a la que pertenece el segmento. El investigador eventualmente tiene que identificar la «reserva de significados», es decir, la categoría a la pertenece cada segmento. ¿Cómo desarrolla un investigador un sistema de organización?

Paso 1: Obtenga una visión del conjunto. Lea el primer conjunto de datos cuidadosamente y continúe leyendo cada conjunto de datos. Anote sus ideas sobre los datos a medida que va leyendo. Esto le proporcionará ideas sobre datos individuales y sobre el fenómeno de interés más amplio. Cuando haya leído al menos tres grupos de datos, pase al paso 2.

Paso 2: Temas generales a partir de los datos. Escoja cualquier grupo de datos para empezar y anote los temas a medida que va leyendo la observación de campo entera o la entrevista. Pregúntese a sí mismo: «¿de qué trata?», «¿sobre qué estuvieron hablando estas personas?», «¿qué era importante en aquel lugar y en aquel momento sobre este pequeño dato?». Un tema es el nombre descriptivo para denominar la materia del segmento. En este momento, usted no se pregunta a propósito de qué están hablando, cosa que identifica el significado del segmento. Cada tema se escribe en el margen de una copia del conjunto de datos. En esta ocasión, no tiene por qué ser exhaustivo con respecto a los temas de cada conjunto de datos. Siga anotando los temas en los márgenes de tres a cinco conjuntos de datos. Vea el extracto 12.2 como ilustración de los temas iniciales en las partes de tres transcripciones. No olvide que algunos temas, como los éxitos escolares o los castigos, son recurrentes.

Paso 3: Compare para la duplicación de temas. Elabore una lista de temas en una cuartilla aparte, con una columna para cada conjunto de datos. Compare los temas en cuanto a la duplicación o solapado de significados. Vea la figura 12.3 para obtener una imagen visual del proceso. Dibuje líneas entre las columnas para conectar temas similares o subraye con colores. Si aparece un tema cuyo significado ya no reconoce, vuelva a leer los conjuntos de datos originales y asegúrese de que lo comprende. En otra hoja de papel, agrupe los temas semejantes: los que aparecen conectados con líneas. Con una lista de todos los temas, agrupe los temas que sean similares, volviendo a los datos si le resulta necesario. Para cada conjunto de temas, el mejor nombre puede aparecer entre las etiquetas de los temas originales o, también, puede crearse otro nuevo que capte mejor la

EXTRACTO 12.2**TEMAS INICIALES EN TRANSCRIPCIONES DE PRÁCTICAS DE DIRECTORES DE PRIMARIA SOBRE EL GRADO DE CASTIGO****Director 1**

- Castigo { «Bueno, como dije, el castigo, es un fracaso, y los fracasos son malos para los niños.
- Éxito escolar { En la educación, nuestra tarea consiste en ayudar a los niños a que se den cuenta de que hay cosas que se les dan bien, y de que la escuela es un lugar para desarrollar sus cualidades. La escuela debería ser un lugar que ayudase a los estudiantes a desarrollar la confianza que tienen en sí mismos.
- Castigo { El castigo no hace nada por promover estas actitudes y, si de mí dependiera, sólo se usaría en casos excepcionales o en ninguna ocasión».

Director 2

- Éxito escolar { «No estamos aquí para hacer que los niños odien la escuela. No estamos aquí para hacer creer a los niños que no pueden aprender. ¿Qué tiene de bueno el castigo? Nuestra política se basa en que
- Normas { el castigo es algo negativo, y tendemos a buscar otras alternativas para ayudar al niño y recurrimos al castigo sólo como último recurso».

Director 3

- Ausencia { «Algo con lo que debemos estar de acuerdo es que... cuando un niño no ha asistido a clase un cierto número de días, sin tener en cuenta las circunstancias, tenemos que considerar la posibilidad del castigo debido simplemente al hecho de que se ha perdido mucha información... mmmm... No podemos hacer nada al respecto. Así pues, en tales casos, el castigo no depende de nosotros, mmm, y hacemos lo que podemos, pero a veces tenemos que castigar al niño sólo por esta razón.
- Alternativa { Mmmm..., es posible que exista una situación en la que, si un padre está dispuesto a trabajar con nosotros, se pueda evitar el castigo: asistir a una escuela de verano, recurrir a profesores, contratar a un profesor particular o llevar al niño a algún programa de verano. Entonces haríamos pasar una prueba al niño cuando regresara y, probablemente, pudiéramos seguir con él... mmm...
- Normas { No está escrito en ningún sitio; pero, según nuestras normas, nos vemos obligados a recurrir al castigo si no se asiste a clase durante un determinado número de días, a menos que podamos demostrar que el niño ha llegado a tener un cierto grado de dominio en la materia de manera que pueda seguir el curso de forma natural».

Temas: castigo, éxito escolar, normas, ausencia, alternativa.

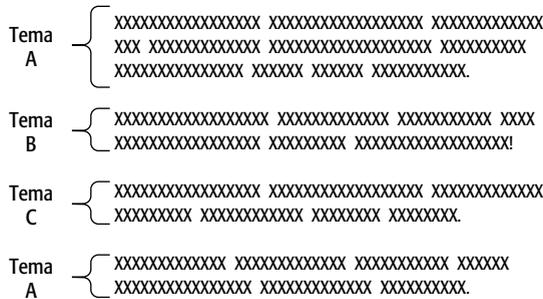
Fuente: Reed, J. S. (1991), *Ethnographic study of the practice of grade retention in elementary schools*. Manuscrito no publicado.

esencia del tema. En este momento, algunos investigadores escriben descripciones provisionales de cada tema. Haga una nueva lista en otra hoja con tres columnas. La primera contendrá los *temas principales* encontrados en los conjuntos de datos; la segunda agrupará los *temas originales* que parezcan importantes para el objetivo de la investigación a pesar de su escasa aparición; en la tercera columna aparecerán los *temas restantes*. Estos temas restantes podrán convertirse en temas importantes o en lo contrario de acuerdo con la formulación de datos avance. En la figura 12.3, conviene destacar que los temas D y E se mencionan

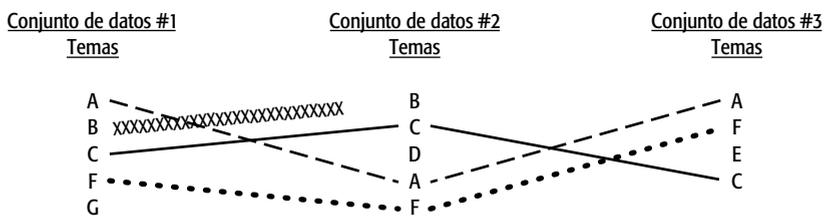
FIGURA 12.3: DESARROLLAR UN SISTEMA DE CLASIFICACIÓN

Paso 1: Identificar los segmentos en un conjunto de datos

Conjunto de datos #1



Paso 2: Anotar los temas dentro de cada conjunto de datos



Paso 3: Clasificar los temas en principales, importantes y restantes

<u>Temas principales</u>	<u>Temas importantes</u>	<u>Temas restantes</u>
A	B	E
C	D	G
F		

Paso 4: Aplicar el sistema de clasificación y continuar perfeccionando

tan sólo una vez en los tres conjuntos de datos, pero cada tema se clasifica indistintamente según su importancia.

Paso 4: Pruebe su sistema de clasificación provisional. Utilizando copias no marcadas de cada conjunto de datos con las que ha trabajado hasta el momento, aplique el sistema de clasificación preliminar, empleando los temas que se encuentran en la primera y en la segunda columna. Escriba un código, una abreviatura del tema, al lado de cada segmento de datos correspondiente. Algunos segmentos pueden tener dos o tres códigos en este momento. Vea el extracto 12.3 como ilustración de las transcripciones que han vuelto a ser codificadas. No hay que olvidar que algunos temas, como los éxitos escolares y el castigo, ahora están formados por subcomponentes y pueden convertirse en categorías. Los éxitos escolares tienen los códigos E cual, E auto y E act; el castigo, tiene los códigos C uso, C alt y C aus. El único tema del principio da lugar a dos temas, normas de la escuela y normas escritas de la escuela, porque el contenido de cada una de ellas es diferente. Este proceso sirve para alcanzar dos

EXTRACTO 12.3

TRANSCRIPCIONES CODIFICADAS SOBRE LOS CASTIGOS EN PRIMARIA

	Director 1
CF	«Bueno, como dije, el castigo, es un fracaso, y los fracasos son malos para los niños. En la educación, nuestra tarea consiste en ayudar a los niños a que se den cuenta de que hay cosas que se les dan bien, y de que la escuela es un lugar para desarrollar sus cualidades. La escuela debería ser un lugar que ayudase a los estudiantes a desarrollar la confianza que tienen en sí mismos. El castigo no hace nada por promover estas cosas y, si de mí dependiera, sólo se emplearía en casos excepcionales o en ninguna ocasión».
E cual	
E auto	
C uso	
	Director 2
E act	«No estamos aquí para hacer que los niños odien la escuela. No estamos aquí para hacer creer a los niños que no pueden aprender. ¿Qué tiene de bueno el castigo? Nuestra política se basa en que el castigo es algo negativo, y tendemos a buscar otras alternativas para ayudar al niño y recurrimos al castigo sólo como último recurso».
E aut	
N esc	
C alt	
C uso	
	Director 3
C aus	«Algo con lo que debemos estar de acuerdo es que... cuando un niño no ha asistido a clase un cierto número de días, sin tener en cuenta las circunstancias, tenemos que considerar la posibilidad del castigo debido simplemente al hecho de que ha perdido mucha información... mmmm... No podemos hacer nada al respecto.

C alt	E alt E alt	Así pues, en tales casos, el castigo no depende de nosotros, mmm, y hacemos lo que podemos, pero a veces tenemos que castigar al niño sólo por esta razón. Mmmm... es posible que exista una situación en la que si un padre está dispuesto a trabajar con nosotros, se pueda evitar el castigo: asistir a una escuela de verano, recurrir a profesores, contratar a un profesor particular o llevar al niño a algún programa de verano.
N escE	E aut	Entonces haríamos pasar una prueba al niño cuando regresara y, probablemente, pudiéramos seguir con él... Mmmm... No está escrito en ningún sitio; pero, según nuestras normas, nos vemos obligado a recurrir al castigo si no se asiste a clase un determinado número de días, a menos que podamos demostrar que el niño ha llegado a tener un cierto grado de dominio en la materia de manera que pueda seguir el curso de forma natural».

Códigos:

- CF = Castigo como fallo
- E cual = Éxitos escolares como cualidades
- E auto = Éxitos escolares como autoestima
- C uso = Uso del castigo
- E act = Éxitos escolares como actitudes
- N esc = Normas de la escuela
- C alt = Castigo alternativo
- C aus = Castigo por ausencia
- E alt = Escuela alternativa
- E P = Elección de los padres
- N escE = Normas escritas de la escuela

Fuente: Reed, J. S. (1991), *Ethnographic study of the practice of grade retention in elementary schools*. Manuscrito no publicado.

objetivos: indica si corresponde bien el nombre del tema con los datos y señala si hay otros temas en los datos que sean importantes pero que no han podido ser codificados con el sistema de organización inicial.

Paso 5: Perfeccione su sistema de organización. Escriba los temas que usted recuerda que aparecen en todos o en casi todos los conjuntos de datos, en una lista; luego realice una lista de temas originales que son importantes para el objetivo de su investigación. A continuación, observe sus temas desde distintos puntos de vista. ¿Hay temas que se acercan más a otros por su contenido? ¿Los hay que son subtemas de otros? ¿Hay otros temas en los datos que no ha reconocido? Hay que perfeccionar la lista de temas constantemente mientras se recogen los datos.

¿Cuántos temas son esenciales? Cuando los investigadores buscan más tarde modelos en sus datos, tienen que recordar todos los códigos de su sistema. Es casi imposible recordar 40 temas distintos. El número de temas depende del objetivo de la investigación y de la naturaleza de los datos. De 25 a 35 temas pueden manipularse con un pequeño esfuerzo; si agrupamos los temas en categorías con subcategorías es más fácil aplicar el sistema de organización.

Desarrollar temas como categorías

Convertir los temas en categorías relativamente sensatas con subcategorías, obliga a los investigadores a pensar de forma más abstracta. Una categoría es un concepto abstracto que representa el *significado* de temas similares. No obstante, los temas pueden formar parte de más de una categoría. Esto es posible porque el contenido de un tema (exposición explícita) puede poseer varias connotaciones (significados implícitos).

Desarrollar las categorías a partir de temas requiere que los investigadores estudien los datos desde diferentes puntos de vista. Los investigadores necesitan evitar maneras de pensar estandarizadas sobre el fenómeno, es decir, rechazar los prejuicios formulados por las personas observadas, investigar sobre lo que realmente quieren decir y explorar todos los aspectos posibles de una categoría. A continuación encontramos algunas estrategias a las que recurren los investigadores (Strauss y Corbin, 1998).

1. *Formule preguntas básicas que desembocarán en preguntas más concretas.* Las preguntas básicas son: ¿quién?, ¿cuándo?, ¿dónde?, ¿qué?, ¿cómo?, ¿cuánto? y ¿por qué? Dar respuestas provisionales a estas preguntas para cada categoría, obliga al investigador a pensar con profundidad analítica.
2. *Analice una oración, una frase o, incluso, una única palabra* que parezca significativa o de interés. Elabore una lista con todos los significados posibles desde el más probable hasta el más improbable. No olvide que una palabra puede tener muchos significados distintos. Estos pueden aceptarse o rechazarse en los datos. Cada significado distinto puede ser una subcategoría.
3. *Compare los datos con una situación similar o diferente.* Al imaginar una situación muy similar, los investigadores pueden distinguir sus características en sus datos y los que se pueden convertir en futuras subcategorías. La misma técnica puede aplicarse a una situación diferente, una situación muy distinta. Esta comparación imaginativa es únicamente una técnica para evitar «no ver nada nuevo» en los datos y para continuar teniendo como objetivo el desarrollo de la consistencia de las categorías.
4. *Identifique las «banderas rojas».* Esta técnica ayuda a los investigadores a ver más allá de lo obvio en los datos, formulando preguntas a partir de los prejuicios de la gente. Algunos ejemplos de «bandera roja» son conceptos o fra-

ses como «nunca», «siempre», «no hay posibilidad de que sea así», «todos saben que se debe hacer así», «no es preciso discutir», etc. Estas frases son señales a las que se les debe prestar mucha atención y formular más preguntas al respecto. Analíticamente, los investigadores nunca dan nada por sentado.

Categorías predeterminadas

Algunos investigadores empiezan con categorías predeterminadas. Hay que recordar que las fuentes de categorías predeterminadas son (a) las preguntas de la investigación, (b) el instrumento de la investigación, como una guía para la entrevista, (c) la experiencia personal que resulta relevante en el campo de trabajo, y (d) las categorías encontradas en la bibliografía. Por ejemplo, la mayor parte de los estudios cualitativos contienen temas, definiciones de los participantes sobre la situación, perspectivas de los participantes, procesos, actividades, acontecimientos, consecuencias, tácticas de participantes, relaciones sociales, estructura social y metodología. Cada categoría se dividiría, luego, en subcategorías a medida que los datos van siendo analizados. Por ejemplo, una categoría «perspectivas de los participantes» puede contener subcategorías para cada perspectiva diferente. Sin embargo, las categorías predeterminadas son sólo puntos de partida. Deben ser aplicadas provisionalmente y pulidas.

Categorías emic y etic

Los investigadores utilizan las categorías emic y etic. Las **categorías emic** representan perspectivas del interior, como por ejemplo: términos, acciones y explicaciones que son distintas del lugar o de la gente. Las **categorías etic** representan las perspectivas del exterior de la situación: los conceptos del investigador y las explicaciones científicas².

Los investigadores cualitativos tienden a destacar los temas emic y las categorías en la formulación de datos porque el objetivo consiste, normalmente, en representar la situación desde la perspectiva de la gente. Los nombres preferidos para los temas y las categorías son los que proceden de los datos. Las categorías emic son explicaciones de lo que el fenómeno significa para los participantes. El extracto 12.4 ilustra cómo los alumnos de segundo curso describen lo que hacen en vacaciones. Destaquemos que los temas «juegos», «vaguear por ahí» y «trastadas» son las palabras empleadas por los niños (lenguaje) cuando describen las actividades de las vacaciones al investigador. «Vacaciones», término empleado en la escuela, es una categoría que tiene tres subcategorías.

² Ver Pelto y Pelto, 1978, para planteamientos emic y etic en Antropología.

EXTRACTO 12.4**CATEGORÍAS EMIC: (ACTIVIDADES DE VACACIONES)**

Juegos	Vaguear por ahí	Trastadas
Un, dos, tres, Perico, Juanico y Andrés	Correr por ahí	Acorralar a alguien
Hotbox	Dibujar sobre los árboles	Poner la zancadilla a alguien
Carreras con relevos	Chupar carámbanos	Mirar bajo la falda de una chica
Carreras	Luchar	Salpicar a alguien
Fuego rojo	Hablar	Tirar bolas de nieve a la cabeza de alguien
Dar patadas al bote	Divertir a la gente	Pegar a alguien y esconderse
Cazar	Dar volteretas sobre la nieve	Robar una gorra de punto.
Platillo	Saltar sobre montones de nieve	
Las chicas pillan a los chicos	Bajar deslizándose por las cuevas de nieve	
Pilla pilla (2)	Bajar deslizándose sobre el hielo	
Avalancha		
Pilla pilla (1)		
Tirar el suéter a la pared		

Fuente: Parrott, S. (1972), «Games children play: Ethnography of a second-grade recess». En Spradley, J., y McCurdy, D. (eds.), *The Cultural Experience*, (Chicago: Science Research Investigation.)

Las categorías etic representan lo que el fenómeno significa para el investigador. Dichas categorías provienen de las experiencias personales del investigador o de su disciplina académica, o se toman prestadas de la bibliografía científica social. Muchos investigadores reprenden el empleo de conceptos científicos sociales durante las primeras fases del análisis de datos porque tales términos sugieren a menudo connotaciones a los que los participantes no han hecho referencia. No obstante, las categorías etic en las últimas fases de análisis de datos resultan esenciales para hacer una situación social distinta comprensible para otros investigadores. No hay que olvidar que en el extracto 12.5, el investigador utiliza, deliberadamente, categorías etic para las actividades de las clases de tercer curso. Aunque los niños hablan de «hacer» diferentes actividades y de «jugar», no dicen de por sí «trabajar» o «juegos». Estas categorías se crearon a partir del lenguaje del investigador. De las nueve categorías de este estudio, tres tenían subcategorías.

LOS MODELOS

El objetivo fundamental de la investigación cualitativa es realizar afirmaciones generales sobre las relaciones entre las categorías al ir descubriendo

EXTRACTO 12.5**CATEGORÍAS ETIC: (ACTIVIDADES O ACCIONES)**

- | | |
|--|--|
| <p>A. Trabajo:</p> <ol style="list-style-type: none"> 1. Hacer Naturales 2. Hacer Sociales 3. Leer 4. Hacer matemáticas 5. Practicar música 6. Deletrear 7. Corregir 8. Redactar 9. Pasar una prueba rápida | <p>C. Jugar con juegos</p> <ul style="list-style-type: none"> Jugar al lavado de cerebro Jugar al verdugo Jugar al holandés volador Jugar a dar patadas al balón Jugar al baloncesto Jugar al perro receptor Jugar a los cambiantes |
| <p>B. Hacer actividades extra:</p> <ol style="list-style-type: none"> 1. Realizar objetos artísticos 2. Confeccionar álbumes de recortes 3. Pintar 4. Hacer cosas para la habitación 5. Escribir poemas 6. Escribir historias | <p>D. Comer</p> <p>E. Asearse</p> <p>F. Hacer gimnasia</p> <p>G. Salir y hablar</p> <p>H. Escoger a nuevos cabecillas</p> <p>I. Leer</p> |

Fuente: Doyle, J. (1972), «Helpers, officers and lunchers: Ethnography of a third grade class». In Spradley, J. y McCurdy, D. (eds), *The Cultural Experience*. (Chicago: Science Research Investigation.)

modelos en los datos. Un modelo es una relación entre las categorías. Conforme los investigadores desarrollan categorías, buscan modelos o relaciones entre ellas. *La búsqueda de modelos significa examinar los datos desde el mayor número posible de puntos de vista.* En la búsqueda de modelos, los investigadores intentan comprender los lazos complejos que existen entre varios aspectos de las situaciones de las personas, sus procesos mentales, sus creencias y sus actuaciones.

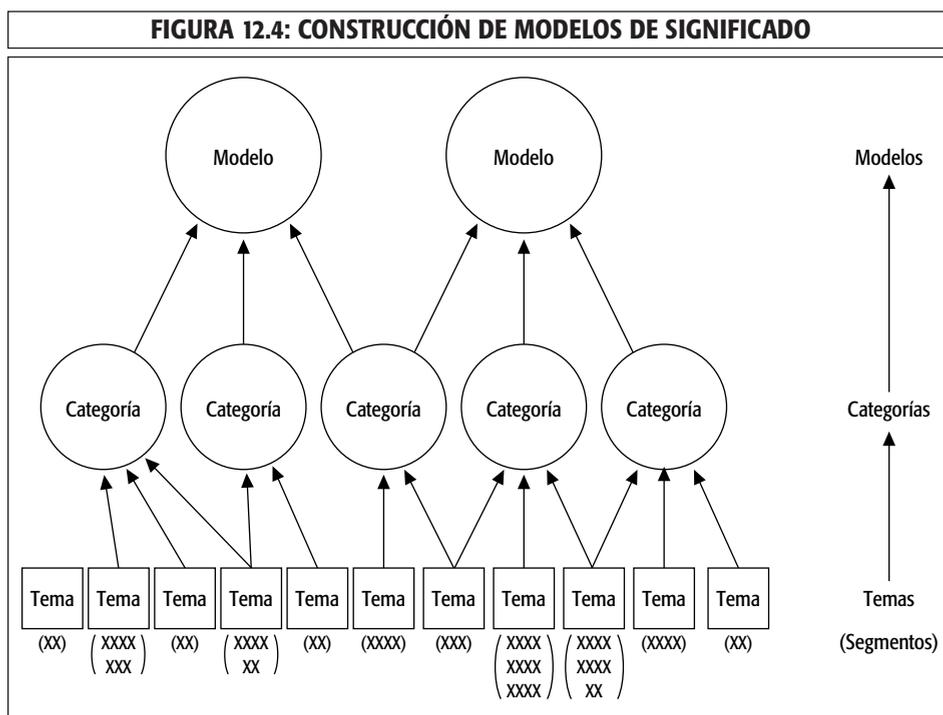
Los investigadores se preguntan cómo afectan las categorías y se ven afectadas por otras categorías. El análisis cualitativo depende de las categorías principales de significados que se extraen de los participantes en la situación.

El proceso de búsqueda de modelos

La búsqueda de modelos empieza a partir de las intuiciones del investigador sobre las relaciones existentes entre los datos. Esto requiere una búsqueda de principio a fin de los datos, comprobando cada intuición a través de la búsqueda de evidencias negativas y explicaciones alternativas. Los investigadores pasan,

luego, a un modo de pensar deductivo –volviendo a repasar todos los temas, categorías y modelos tanteados para buscar confirmación–. Una parte de esta fase consiste en evaluar si la información de los datos es adecuada, inútil y centrada. Los investigadores determinan si los datos iluminan suficientemente bien el problema de la investigación y qué datos son los esenciales para la historia que se está descubriendo a partir de los fenómenos seleccionados.

Los modelos pueden adoptar formas distintas y niveles de abstracción que dependen del objetivo y empleo del estudio. Los modelos también están relacionados con la estructura conceptual seleccionada para la encuesta. Los modelos más importantes sirven tanto como las estructuras a la hora de presentar los descubrimientos y organizar los informes. El proceso queda esquemáticamente recogido en la figura 12.4. Destáquese que el número de segmentos (las x) que representa un tema, varía. Algunos temas se pueden incluir dentro de más de una categoría y otros temas no son esenciales para el problema de la investigación. Y más aun, una categoría puede colocarse dentro de más de un modelo. Esta elasticidad de los significados de tema y de categoría permite que surjan los *modelos de significados*. Los significados del tema, categoría y modelos dependen del



Fuente: adaptado de A. Vierra y J. Pollock (1992), *Reading educational research*, (2ª edición), Scottsdale, AZ: Gorsuch Scarisbrick Pub, p. 262.

contenido de cada uno y de la comparación hecha con dicho contenido: otros temas, otras categorías y otros modelos.

Supongamos que uno identifica el tema «pizza». Su descripción hablaría probablemente de la masa y de los ingredientes como distintivos de los temas «hamburguesa» y «leche». Los temas de pizza y hamburguesa quedan agrupados dentro de una categoría denominada «comida rápida», que se distingue de las categorías «cena para cuatro» y «picoteo». Luego empezamos a formular otras preguntas para el estudio –modelos nutricionales sobre las costumbres alimenticias de los miembros de la familia–. Parte de cada categoría podría incluirse en un modelo de comida nutricional, y parte de cada categoría podría incluirse en un modelo de comida no-nutricional. Señalemos que, en cada nivel de abstracción, el significado deriva parcialmente de la categoría a la que pertenece. Desde luego, también podemos buscar otros modelos entre los temas y las categorías, como opiniones sobre la comida y la nutrición, sobre el grado de conocimiento acerca de la nutrición y el carácter social de la comida. De esta manera, los investigadores cualitativos construyen cada nivel de abstracción, incluso, si el proceso parte de intuiciones, del problema de la investigación y de la estructura conceptual seleccionada.

El proceso es, normalmente, un proceso circular en el que se vuelve siempre a los datos para validar cada modelo y luego modificar o dejar la idea como parte de una abstracción más amplia. Aunque algunos procesos son tediosos y llevan mucho tiempo, también se trata de «un proceso creativo que requiere realizar cuidadosamente juicios sobre lo que realmente es importante y tiene significado en los datos» (Patton, 1990, p. 406).

Las técnicas de la búsqueda de modelos

Las siguientes técnicas son estrategias que facilitan la búsqueda de modelos y que no tienen la finalidad de ser exhaustivos. Cada investigador debe seleccionar las estrategias que aclaran los modelos en sus datos. El rigor, la duración y el proceso del análisis de datos variará; los investigadores cualitativos están obligados a dirigir y registrar sus propias técnicas analíticas y procesos de la manera más completa posible.

Estimar el carácter fidedigno de los datos. Aunque el carácter fidedigno de los datos se estima en el momento de cada experiencia de campo y en la recopilación de datos, también es importante durante el análisis de datos. El investigador selecciona evidencias fidedignas para la búsqueda de modelos a través de datos solicitados o no solicitados para la valoración cualitativa, influencias sutiles entre las personas del lugar, afirmaciones imprecisas o específicas y la exactitud de las fuentes (¿una persona observadora?, ¿una persona reflexiva?, ¿una persona emocional?, ¿una persona subjetiva?). Seleccionar los datos fidedignos también implica un conocimiento de las conjeturas por parte del investigador, predisposiciones e influencia en la situación social.

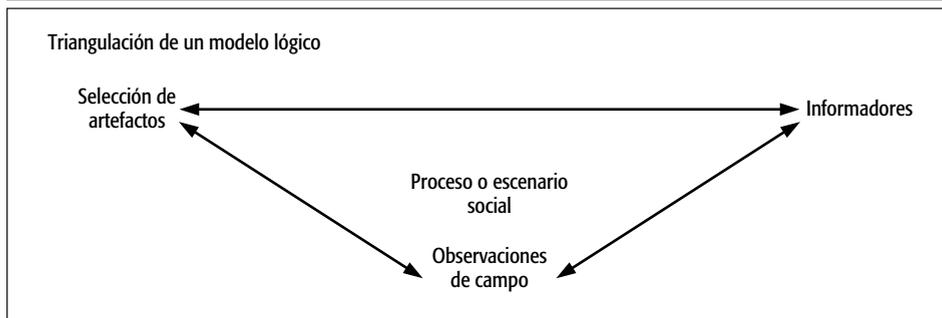
Usar la triangulación. Los investigadores utilizan la triangulación (Denzin, 1978), que es la validación entre las fuentes de datos, las estrategias para la formulación de datos, los periodos temporales y los esquemas teóricos (ver capítulo 10). Para encontrar regularidades en los datos, el investigador compara las diferentes fuentes, situaciones y métodos para ver si se repite el mismo modelo. Un tema de «colaboración institucional», por ejemplo, puede comprobarse si comparamos los datos encontrados en la recopilación de artefactos (apuntes, informes, folletos oficiales, cartas), entrevistas (proyectos de codirectores, profesores, directores), y observaciones de campo de reuniones del proyecto. La figura 12.5 ilustra este método de la triangulación.

No obstante, los investigadores sienten que, a pesar de que ellos sólo observan, oyen o graban directamente una frase, para algunos tipos de análisis, un solo incidente está lleno de significado.

Evaluar las evidencias discrepantes o negativas

Los investigadores llevan a cabo una búsqueda activa de evidencias discrepantes o negativas que modifiquen o rebatan un modelo (ver capítulo 10). Un caso negativo es una situación, escena social o punto de vista del participante que contradice un modelo de significados. En otras palabras, hay dos modelos más que uno. Los datos discrepantes presentan una variante de un modelo. Estas excepciones son muy útiles porque hacen que el modelo original sea más característico y proporcione nuevas percepciones para modificar los modelos. Por ejemplo, un modelo puede ser que una acción particular tenga lugar en la mayor parte de las situaciones excepto para este tipo de situación³.

FIGURA 12.5: TRIANGULACIÓN. (TRIANGULACIÓN DE UN MODELO LÓGICO)



³ Ver Strauss, A y Corbin, J. (1998), *Basics of qualitative research: Grounded theory procedures techniques, for a conditional matrix*.

Ordenar categorías para los modelos

Podemos ordenar categorías de varias maneras para descubrir modelos. Una manera es ordenar las categorías según como sucedieron. Los investigadores se preguntan: ¿cuál es la situación o acción que ocurre primero?, ¿el acontecimiento está acompañado por una o más opiniones?, ¿cuál fue la consecuencia? Ordenar las categorías en una secuencia resulta útil durante un proceso de análisis para identificar cambios producidos de un momento a otro (ver extracto 12.6).

Una segunda manera es aumentar, combinar, incluir y crear nuevas categorías que tengan sentido empírico y lógico, que compartan el significado. El «sentido lógico» de la búsqueda de modelos es que el significado de una categoría se

EXTRACTO 12.6

ORDENAR CATEGORÍAS PARA LOS MODELOS

Consecuencias

El comportamiento de los niños ante el ritmo se desarrolla de manera jerárquica. Los niños modifican considerablemente sus manifestaciones de acciones concretas o estrategias. Sin embargo, [en] el avance de los cambios de comportamiento, exhibidos por los niños observados... cada sesión incorporaba lo que se había obtenido en la sesión precedente como lo que era nuevo en las conductas de los niños. Cada sesión mostraba un objetivo cualitativamente diferente de las actividades y experiencias de los niños...

Al mantener esta perspectiva, el comportamiento ante el ritmo observado en esta investigación puede representarse en dos fases. Durante la primera, los niños adquieren y desarrollan cualidades y comportamientos que se necesitan para cambiar de forma estándar. Se centran en comprender el movimiento rítmico, desarrollar los movimientos corporales necesarios para adaptarse al ritmo y coordinar estos movimientos corporales con el ritmo. El comportamiento exhibido durante la primera fase puede describirse como el comportamiento que se desarrolla ante el ritmo...

Durante la segunda fase, los niños casi desarrollan comportamientos básicos y son capaces de centrar su atención en comparaciones con sus amigos y exploraciones de su propio potencial y del ritmo. El refinamiento, la extensión y la elaboración del comportamiento rítmico de los niños son evidentes durante esta fase. Este comportamiento puede describirse como el comportamiento desarrollado ante el ritmo...

Conclusiones

Descrito, ya sea por fases o por sesiones, el movimiento rítmico surge en un orden predecible...

Fuente: Fox, J. E. y Tipps, R. S. (1995), «Young childrens' development of swinging behaviors», *Early Childhood Research Quarterly*, 10,491-504.

ve influenciado por su relación con un modelo. En nuestro ejemplo de categorías de comida nutricional y no-nutricional, dichas categorías pueden mostrar un modelo de la manera de comer del entorno social de los adolescentes (familia cercana, amigos del instituto, abuelos, otros adultos, niños más pequeños o ellos mismos). Al crear una nueva categoría de valor nutricional (comida equilibrada) que incluye comida nutricional y no-nutricional, se debe buscar un modelo de elección de comida: ¿qué comida escogen nuestros hijos adolescentes en diferentes situaciones sociales?

Clasificación de las categorías para los modelos. Los investigadores agrupan las categorías de varias maneras para identificar los significados. En un estudio sobre directores y profesores poco satisfactorios, la categoría de profesores insatisfactorios se colocó en primer lugar por tipos de profesores insatisfactorios y, luego, por tipos de respuestas (Luck, 1985). Cada categoría fue reorganizada para ver si había un modelo entre los tipos de profesores insatisfactorios y los tipos de respuestas. Cuando no se encontró ningún modelo, otra categoría –métodos de identificación de profesores insatisfactorios– se clasificó con los tipos de respuestas. Esta clasificación condujo a un modelo que contemplaba desde los métodos de identificación profesores insatisfactorios hasta los tipos de respuestas.

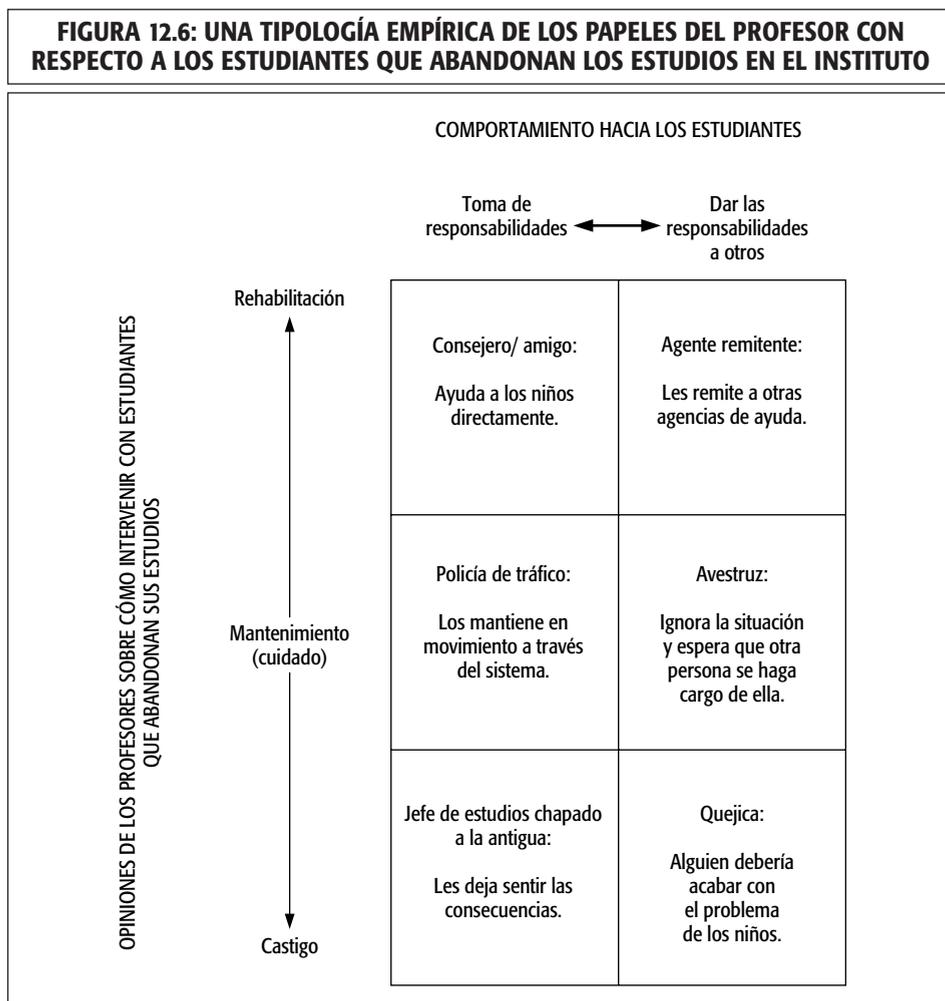
Construcción de representaciones visuales. Los investigadores construyen representaciones visuales, un conjunto organizado de información, como figuras, cuadros, diagramas integradores y gráficos, que ayudan en el análisis. Hay muchas formas de representación visual. La figura 12.6 y el extracto 12.8 son posibles ejemplos. La mayor parte de los investigadores que utilizan un acercamiento descubridor, cuidan no llevar a cabo un cierre precipitado de la construcción de diagramas integradores. Los datos descriptivos contextuales deben acompañar los diagramas, una abstracción que limita los tipos de interpretación que se pueden dibujar. Las representaciones visuales son instrumentos y no son «realidad» por sí mismos. Los diagramas ayudan a los investigadores a acercarse a un análisis más abstracto permitiéndoles plantearse diferentes preguntas sobre los datos. Los investigadores tratan de equilibrar el respeto a la complejidad de la realidad con la necesidad de simplificar para conseguir objetivos analíticos y comunicativos. Una vez finalizados los diagramas integradores, éstos tienen la función de representar de forma visual el estudio entero y se adjuntan al informe.

Realización de análisis-cruzados lógicos. Presentados normalmente en formato de cuadros, las categorías se cruzan entre sí para originar nuevas intuiciones para nuevos análisis de datos. Estas categorías-cruzadas revelan discrepancias lógicas en los datos ya analizados y sugieren áreas en las que los datos y los modelos pueden quedar lógicamente encubiertos. No obstante, el investigador *no* debe permitir que estos cuadros dominen el análisis, sino que debe utilizarlos para generar inspiraciones en la búsqueda de modelos. En la figura 12.6 las seis categorías de funciones del profesor con respecto a estudiantes que abandonan los

estudios en el instituto, primero se desarrollaron y luego se sometieron a un análisis-cruzado, que permitió descubrir dos dimensiones de significado: los comportamientos y las opiniones.

Credibilidad de las explicaciones

Conforme los investigadores van construyendo sus categorías y buscando modelos, necesitan encontrar otras explicaciones verosímiles para justificar los lazos entre las distintas categorías. Siempre hay explicaciones alternativas, pero



Fuente: de M. Q. Patton (1990), *Qualitative Evaluation and research methods*, 2ª edición, Newbury Park, CA: Sage, p. 413.

puede que no queden razonablemente respaldadas por los datos. *Un modelo se convierte en una explicación sólo cuando los modelos alternativos no ofrecen explicaciones razonables sobre el problema de la investigación.* La credibilidad es una manera de comprobar la calidad de los datos dentro de las limitaciones diseñadas. La credibilidad se demuestra a través de la presentación de datos y el rigor del análisis.

La siguiente analogía ilustra un caso de credibilidad. Si una persona pela una manzana, en primer lugar tendría conciencia de la manzana entera –su tamaño, su forma y otras características–. Inicialmente, se entra en contacto con la piel; la piel también tiene características distintivas. Bajo la piel se encuentra la pulpa. Mucha gente, en este momento, deja de pelar la manzana y quedan satisfechos con su conocimiento sobre la manzana. Otros son más curiosos sobre por qué hay tanta pulpa. Continúan indagando y, finalmente, descubren una membrana incrustada en la pulpa. Algunos se preguntan por qué hay una membrana. Al cortar cuidadosamente la membrana, obtienen una semilla. Llegados a este punto, por fin se comprende: todas las piezas «encajan» entre sí formando un conjunto. Hay una explicación de por qué la membrana, la pulpa y la piel protegen el corazón de la manzana, la semilla. En cada fase de corte de la manzana, había explicaciones alternativas para entender las partes de una manzana; sin embargo, con el último descubrimiento (la semilla), no hay otra explicación razonable para el significado «real» de las partes de una manzana. Destaquemos que para descubrir el significado de una manzana, uno debe continuar formulándose preguntas e investigando sobre cada porción hasta alcanzar el corazón de la manzana.

MANUAL DE TÉCNICAS INFORMÁTICAS EN EL MANEJO DE DATOS

Los estudios cualitativos destacan por contener «montones» de datos. Los investigadores tienen que manejar dichos datos con el fin de realizar análisis y de describir los objetivos. El manejo de datos consiste en utilizar un sistema para recuperar los grupos de datos (notas de campo y transcripciones de entrevistas) y para reunir los datos codificados en un lugar. Los investigadores, generalmente, emplean uno de los tres planteamientos para el manejo de datos: técnicas manuales, programas de software generales que pueden adaptarse a posibles necesidades de investigaciones cualitativas y programas de software especiales diseñados para el manejo de datos cualitativos. Normalmente existen muchos programas de software generales y más de 24 programas de software especiales, varios de los cuales cuentan con versiones perfeccionadas. Dada la confidencialidad de los datos, no se recomienda archivar los datos cualitativos en un ordenador. Primero hablaremos sobre la creación de un sistema de expedientes y, luego, del manejo manual e informático de datos.

Creación de un sistema archivador de datos

Resulta esencial crear un sistema para archivar los datos a medida que empiezan a acumularse. Un sistema para archivar datos es un procedimiento que sirve para identificar y recuperar un grupo particular de notas de campo originales o una transcripción de una entrevista. Todas las páginas de cada tipo de recopilación están numeradas por orden para localizar la información dentro de cada grupo de datos. Además, cada conjunto de notas de campo o grupo de transcripciones queda identificado en la primera página mediante la fecha, lugar, persona(s), lugar exacto y acontecimiento o escenario social. Puesto que trabajar con datos implica una reorganización a mano o por ordenador, se realizan varias copias y se guarda una como copia de seguridad. Esta copia debe permanecer en un lugar en el que no sufra alteraciones, se pierda o se destruya.

Un sistema para archivar datos contiene cada conjunto de datos en cuadernos de notas, carpetas de documentos o en un ordenador personal (PC). El objetivo es el mismo, localizar los datos de campo. El sistema de expedientes depende, parcialmente, del formato de los datos y del objetivo del estudio. Los investigadores archivan datos mediante la utilización de uno o más de los sistemas siguientes:

- por fecha de observación o entrevista, especialmente si lo más importante es el proceso de análisis;
- por lugar, especialmente si el estudio incluye más de un tipo de lugares físicos: casas, escuelas, agencias, oficinas;
- por persona entrevistada, numerados frecuentemente como entrevista 1, entrevista 2, etc, con una lista separada para identificar cada número específicamente;
- por tipo de escenario social, especialmente, si muchos tipos de escenarios sociales tienen lugar para un número pequeño de personas, como clases, conferencias padres-profesores, reuniones, etc.

Planear por adelantado la codificación por colores, numeración y el empleo de abreviaturas es inestimable a la hora de realizar el análisis de datos y la redacción del informe final.

Manual de manejo de datos

Hay varias maneras de manejar datos manualmente después de que los temas hayan sido codificados en una copia de los datos. La codificación final de los temas también se convierte en una copia principal y se guarda en lugar seguro. Los investigadores hacen fotocopias de los datos codificados y trabajan sólo con copias para el análisis intensivo y la redacción del informe.

Hablaremos sobre las dos maneras de manejar los datos codificados manualmente. Una técnica es el planteamiento corta y archiva; la segunda técnica es el sistema de las tarjetas expediente (Bodgan y Biklen, 1998). Existen muchas variedades de estas dos técnicas.

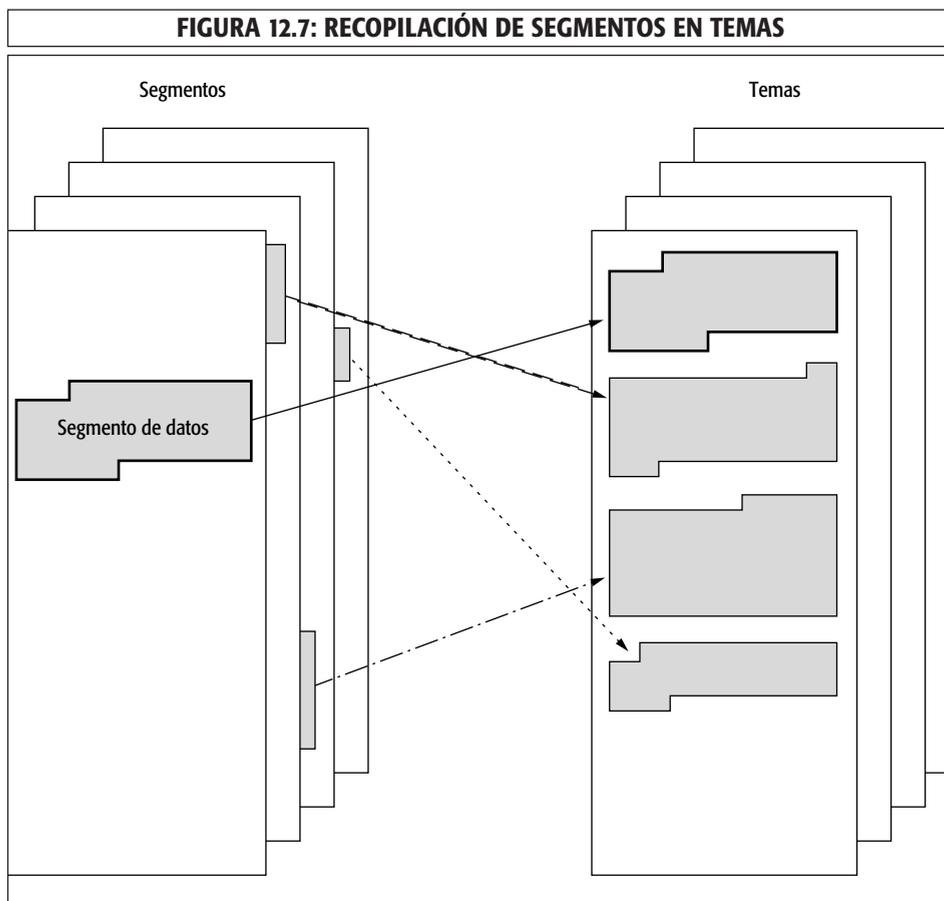
La *técnica corta y archiva* requiere múltiples copias de cada conjunto de datos para cortar. Las carpetas están etiquetadas con el nombre de los temas. Cuando alguien recorre los datos (que pueden ir desde una frase hasta un párrafo o una serie de párrafos), cada tema aparece marcado, primero, para identificar la fuente y el número de la página. Esta clasificación puede realizarse con números o con cualquier otro sistema. Hay que prestar cuidado a que los números de los expedientes hagan referencia a las fuentes, a los números de página y a los temas y no a los entrevistados, ni a los lugares de campo, etc. Sólo después de que cada segmento de tema posea una etiqueta de identificación, deben dividirse las páginas de datos en segmentos y colocarlas en la carpeta adecuada de cada tema. Si algunos segmentos de datos contienen más de un tema, entonces se necesitan copias para cada carpeta.

Una variedad es colocar la página entera del conjunto de datos que contiene el tema marcado en una carpeta. Otra posibilidad es pegar o copiar los datos del tema en tarjetas de 7'62 por 12'70 cm y luego agrupar las tarjetas de cada tema en un montón. Cada tarjeta debe estar marcada por tema y fuente. De ambas maneras, el resultado es una caja de carpetas de expedientes o de tarjetas que contienen todos los segmentos de datos importantes de cada tema. La figura 12.7 representa gráficamente el proceso.

Llegados a este punto, muchos investigadores escriben el contenido de cada carpeta o grupo de tarjetas y seleccionan el material ilustrativo. No importa qué carpeta se selecciona para leer primero. Al colocar todas las carpetas de temas importantes juntas y leerlas, los investigadores transforman los temas en categorías. Si existen muchos segmentos de datos en una carpeta o muchas tarjetas en un montón, el contenido puede ser una categoría que necesita dividirse en subcategorías.

La *técnica de tarjetas* es un sistema de clasificación para localizar los temas dentro de un banco de datos entero. El investigador tiene una fotocopia de los datos con los temas codificados de cada conjunto de datos. El investigador también tiene su lista de categorías para el estudio. Cada conjunto de datos (fuente) debe ir numerado, al igual que las páginas y las líneas. Así pues, cada tema tiene tres métodos de identificación –el documento (o fuente), la página y la/s línea/s–. Los investigadores leen sus datos, seleccionan la categoría de cada tema y escriben en el índice de las tarjetas tres identificaciones: el número de la fuente, el número de la página y el número/s de línea. El resultado de este proceso es un índice de datos relevantes para cada categoría.

Este método tiene algunas ventajas. No tenemos voluminosas carpetas ni requiere recortar. Algunos inconvenientes son (1) que hay que leer toda la página



Fuente: R. Tesch (1990), *Choices among Qualitative Analysis Programs Presented at the American Evaluation Association Annual Meeting*, Washington, D.C.

para localizar el segmento que se busca, (2) que hay que recuperar todos los datos para leerlos cuando están dispersos por todo el banco de datos, y (3) volver a dejar cada página en su sitio dentro del sistema de expedientes para hacerla accesible en la búsqueda de modelos. No obstante, algunos investigadores, que usan cuadernos de notas para almacenar sus datos, prefieren esta técnica.

Usar programas informáticos

Los programas informáticos utilizados para el manejo de datos cualitativos pueden ser desde programas generales producidos por fabricantes de software no especializado, hasta programas especializados, normalmente desarrollados por

investigadores cualitativos. Los programas generales que adoptan los investigadores para manejar los montones de datos van desde procesadores de texto hasta programas especiales como código-y-recuperadores, teoría de la codificación y redes conceptuales (Weitzman y Miles, 1995). Además, el empleo de programas software en investigaciones cualitativas es una innovación técnica que cambia muy rápidamente. Con este texto pretendemos explicarle todas las posibilidades de las que usted dispondría si utilizara programas informáticos para el manejo de datos cualitativos. La tabla 12.1 hace una lista de los usos del software informático en la investigación cualitativa. Primero hablaremos de los programas generales y, luego, de los específicos.

Los programas generales incluyen los procesadores de texto, los recuperadores de texto y los administradores de base de texto. Los *procesadores de texto* están diseñados para la elaboración y revisión del texto y, por lo tanto, son útiles para escribir o editar notas de campo, para transcribir entrevistas, para redactar los apuntes, para preparar expedientes para codificar y analizar, y para escribir informes. La mayor parte de los procesadores de texto buscan cadenas de caracteres, y algunos crean enlaces de hipertexto, que conectan dos puntos en el texto de manera que se pueden ver ambos simultáneamente. Los programas *recuperadores de texto* son programas sofisticados de exploración de texto que se especializan en la búsqueda de todos los casos de palabras, frases (u otras cadenas de caracteres), y sus combinaciones en un expediente o más. Los programas *administradores de base de texto* comparados con los programas recuperadores de texto, ofrece en más

TABLA 12.1: EMPLEO DE ORDENADORES EN LA INVESTIGACIÓN CUALITATIVA

1. Tomar notas en el campo.
2. Escribir o transcribir las notas de campo y los datos de las entrevistas.
3. Editar, corregir y elaborar las notas de campo y las transcripciones de las entrevistas.
4. Codificar y colocar las palabras clave o símbolos para segmentar los datos y posibilitar su posterior búsqueda.
5. Archivar y mantener los datos en una base de datos organizada.
6. Buscar y archivar los segmentos relevantes y los trozos de datos y hacerlos disponibles para que puedan ser examinados.
7. «Enlazar» y conectar los segmentos de datos importantes entre sí, establecer categorías y modelos para un examen más detallado.
8. Escribir comentarios personales sobre algunos aspectos de los datos en forma de una base para análisis más profundos.
9. Representación visual y colocación de los datos seleccionados en un formato denso y organizado como una ficha modelo o una red para su examen; crear diagramas que representen los descubrimientos o las teorías.
10. Preparar informes provisionales, borradores y definitivos.

posibilidades para organizar y hacer subconjuntos del texto sistemáticamente y, además, aseguran su búsqueda y recuperación. Normalmente presentan funciones especializadas para manejar y organizar los datos y, especialmente, para crear diferentes subconjuntos de datos para análisis posteriores.

Existen tres tipos de programas software que tienen como objetivo ayudar en el análisis de datos cualitativos (QDA; *qualitative data análisis*). Dichos programas, aunque, generalmente, se basan en la lógica formal, no desarrollan una teoría del investigador o temas explicativos; los programas de software desconocen el significado de los datos cualitativos. Los programas código-y-recuperadores ayudan al investigador a dividir el texto en segmentos o trozos, y buscan todos los trozos con un código dado o una combinación de códigos. Estos programas reemplazan las tareas manuales de cortar y archivar o efectuar un sistema de tarjetas. Los programas varían según su capacidad de hipertexto. Los programas de teoría de la codificación, normalmente, incluyen las mismas características que los programas código-y-recuperadores; pero, también se utilizan para establecer conexiones entre códigos y desarrollar categorías y modelos. Los programas conceptuales de construcciones de red también ayudan en el establecimiento de la teoría, pero el investigador trabaja con redes gráficas. Las redes son «redes semánticas» desarrolladas a partir de los datos y de los conceptos de los investigadores.

¿Cómo funciona el análisis de datos asistido por ordenador? El proceso es similar a las técnicas del manejo de datos manual que ya hemos descrito. Los investigadores pueden unir símbolos, que ya han determinado, a segmentos de datos, extractos de segmentos y unir estos segmentos en un lugar. Los programas indicarán automáticamente en cada segmento de dónde proviene (página y línea de los datos). Si usted prefiere trabajar con copias en formato papel, el programa imprimirá todos los datos o parte de ellos. Los datos pueden volverse a codificar.

Los programas software para el manejo de datos cualitativos funcionan a partir de MS-Dos, Microsoft Windows y Macintosh. Los estudiantes decidirán, primero, qué tipo de usuarios son, si lo utilizarán para un proyecto o para los siguientes años, qué tipo de base de datos y proyecto, qué tipo de análisis se prevé, qué importancia tiene la confidencialidad de los datos, y el coste límite.

La «confidencialidad» de los datos frente al «distanciamiento» hace referencia a observar sólo pequeñas porciones de los datos al mismo tiempo en comparación con la sensación de inmersión en los datos que proviene de ojear montones de papeles o leer páginas de datos de un expediente al tiempo. Otros dos puntos críticos son el uso de la cordialidad y el programa de la flexibilidad, es decir, la habilidad para modificar los códigos iniciales y para personalizar y adaptar el programa al estilo analítico del investigador y al tipo de base de datos. Por ejemplo, los datos de la observación y de las conversaciones informales están menos estructurados que los datos de las entrevistas estándares de respuestas abiertas. Algunos programas requieren bastantes datos estructurados para ser eficientes.

Hay opiniones diferentes sobre si los investigadores principiantes deberían utilizar programas de software diseñados especialmente para los aspectos mecánicos del análisis de datos. Los desacuerdos se centran en el tiempo que se invierte para aprender a usar un programa y en la cantidad de tiempo que se ahorra. Para las personas que se manejan bien con los ordenadores y a los que se les da bien aprender programas nuevos, el tiempo y el dinero invertidos valen la pena, especialmente en un estudio importante como una disertación o para continuar una carrera en la investigación.

Los investigadores con experiencia también tienen preferencias. Algunos quieren mantener sus cualidades analítico-cognitivas separadas de su manejo de datos. Otros investigadores defienden el manejo de datos informático y el análisis asistido por ordenador.

No obstante, hay diferencias significativas entre los programas que los usuarios deberían conocer antes de elegir. Libros recientes que comparan los programas de software deberían consultarse antes de invertir tiempo y dinero en un programa de software⁴. Además, los actuales programas de software para analizar datos cualitativos (QDA; *qualitative data análisis*) pueden encontrarse en Internet en las páginas Web esponsorizadas por algunos individuos para su empleo en universidades y por organizadores comerciales. *QualPage*, por ejemplo, hace una lista de los recursos para los investigadores cualitativos e incluye programas de software QDA. Otras páginas QDA presentan breves análisis sobre cada programa. Algunos programas software QDA también pueden solicitarse a través de Internet. Por ejemplo, uno de estas páginas proporciona noticias sobre la versión actual y proporciona información con ejemplos sobre cómo utilizar el programa; una página con un breve manual y preguntas frecuentes contestadas; una página fuente con hipervínculos hacia otros sitios relacionados con QDA e investigación científica social; y una página de pedido para poder comprar el software.

VARIACIONES EN ESTRUCTURA NARRATIVA Y REPRESENTACIÓN

Una característica de la mayor parte de las investigaciones cualitativas es una presentación narrativa de los datos y la diversidad de representaciones visuales de los datos. Los datos se presentan como citas del lenguaje de los participantes,

⁴ Ver Fielding, N. G. y Lee, R. M. (1998), *Computer analysis and qualitative research*; Kelle, U. (1995), *Computer-aided qualitative research*; Tesch, R. (1993), «Personal computers in qualitative research», en LeCompte, M. D. y Preissle, M. (1993), *Ethnography and qualitative design in educational research* (2ª ed.), p. 279-314; Weitzman, E. A. y Miles, M. B. (1995), *Computers programs for qualitative data analysis*.

mencionando las notas de campo y las transcripciones de las entrevistas como fuentes. No todos los datos se recopilan en un único estudio, ni tampoco necesariamente en el mismo formato para todos los lectores. Hay cuatro auditorios potenciales: académico, participantes en el estudio, elaboradores de planes o políticas y el público general. La estructura narrativa depende de las complejidades del fenómeno, del objetivo de la investigación, de la tradición de la investigación y del auditorio.

Los autores cualitativos recomiendan una estructura de conjunto que no siga ciegamente la introducción cuantitativa estándar, los métodos, los resultados y el formato de la discusión (Richardson, 1998; Van Manen, 1988). Los métodos pueden denominarse «procedimientos» o «estrategias»; los resultados se designan «descubrimientos». Los subtítulos utilizan, a menudo, el lenguaje expresivo de los participantes más que la retórica científica social. El estilo de redacción puede ser personal, inteligible y asequible a un auditorio amplio. El nivel de los detalles hace que la narración parezca «real» y «viva» introduciendo al lector directamente en el mundo real.

Auditorio y presencia del autor

Hay un número de consecuencias que el investigador debe tener en cuenta cuando redacta su informe: la recopilación para auditorios específicos, las representaciones visuales de los descubrimientos y la presencia del autor (Creswell, 1988). La recopilación es el empleo de los recursos literarios para dar forma a un informe dirigido a un auditorio concreto. Para auditorios académicos y publicaciones, la abundante exposición de títulos académicos del autor, las referencias, las notas a pie de página, las secciones de metodología y el uso de metáforas académicas, resultan apropiados. Para los auditorios morales o políticos resulta conveniente el empleo de palabras con alta connotación «excluyente» como mujer, mujeres, feminista en escritos feministas, la moral o las justificaciones del autor y las metáforas de refuerzo. Para los participantes y los elaboradores de políticas existe menos bibliografía y teoría, una ojeada breve metodológica, y una descripción detallada de la práctica, sin olvidar los asuntos prácticos o los resultados y utilizar las metáforas de «sentido común».

La presencia del autor se reconoce a través de varios recursos intelectuales. El papel del investigador queda descrito. Además, la presencia del autor puede dejarse ver a través de las notas a pie de página, de comentarios de interpretaciones o un epílogo⁵. En algunos estudios críticos, los objetivos políticos del investigador quedan reflejados en las secciones introductorias.

⁵ La presencia de algunos autores se descubre por la firma o el discurso, esto es, el texto posee un ritmo, una cadencia y una expresión que hacen fácilmente identificable el trabajo como perteneciente a un determinado autor (Clandinin y Connelly, 1998).

Los estudios cualitativos presentan una variedad de estructuras explicativas: interpretaciones descriptivo-analíticas, informes realistas, abstractos debates teóricos y difusión de relatos personales. A pesar de esta variación, generalmente, dos aspectos de un estudio se presentan como datos: el contexto y las citas sobre el lenguaje de los participantes. Las manifestaciones de la gente observada y entrevistada representan sus construcciones del mundo, los significados que dan a las situaciones sociales y a sus experiencias. La tarea del investigador es ordenar estos puntos de vista de una manera lógica, y conseguir que lo que los participantes quisieron decir sea comprensible para el lector. En primer lugar, presentaremos lo que es común a la mayoría de los estudios cualitativos y, luego, hablaremos sobre las características diferenciadoras de las estructuras descriptivas y representaciones de cinco planteamientos de investigación cualitativa (ver capítulo 2).

Estructurar la narración y el lenguaje de los participantes

Un estudio puede estructurarse de varias maneras: el contexto natural, la experiencia fenomenológica, las teorías relevantes seleccionadas de las anotaciones de los datos de campo, o una orientación política. El **contexto** de un estudio es la descripción de las situaciones de la gente y de los acontecimientos en los que tiene lugar el fenómeno que nos interesa. Cómo se estructura un estudio varía según las tradiciones cualitativas: etnográficas, fenomenológicas, estudios de caso, teoría basada en hechos o estudios críticos.

En estudios etnográficos el contexto cultural compartido por el grupo es el lugar, los participantes y el periodo de tiempo en el que se efectúa la recopilación de datos. El contexto holístico de la investigación se considera un descubrimiento importante; el espacio se describe y emplea para identificar los elementos contextuales que influyen en la vida cultural o en el «caso». En un estudio sobre el primer año de funcionamiento de una escuela innovadora, por ejemplo, el contexto es el plan y los detalles de la construcción de la nueva escuela, el nuevo claustro al completo y el claustro encargado del curso académico. También forma parte del contexto cómo la escuela innovadora se ajusta a sistemas más amplios: sistema escolar de la comunidad y sistema educativo del estado.

La estructura para el estudio fenomenológico es el tipo de experiencia que ha tenido lugar o, todavía, está ocurriendo. La experiencia debe haberla vivido tanto el investigador como el entrevistado, por ejemplo, ser padre/madre trabajador/a soltero/a, una mujer directora que dimite voluntariamente, o un niño que ha vivido la muerte de un hermano. La situación natural se utiliza en descripciones de «qué» ocurrió y «cómo» tuvo lugar la experiencia.

Mientras que el contexto natural es muy importante en la etnografía y en la investigación del estudio de caso, resulta menos importante en los estudios de

teoría basada en hechos. En la teoría basada en hechos, la identificación de las teorías que constituyen la estructura del estudio es crucial, y una breve descripción del contexto natural (esto es, salas de los hospitales, patios de las escuelas, grupos de apoyo contra el cáncer), forma parte de la metodología. Puesto que hay diferentes tipos de estudios críticos, la narración puede estructurarse a partir del contexto natural, de una estructura teórica o de una estructura política (orientación o punto de vista político del investigador). En la mayor parte de los estudios, la estructuración de la descripción que debe seguirse es esencial para que los lectores puedan llegar a comprender el estudio y añadir los conocimientos adquiridos a la investigación futura o a las prácticas. Todos los nombres propios están codificados por confidencialidad.

La presentación del lenguaje de los participantes es preceptiva, porque *constituyen los propios los datos*. Los datos pueden presentarse en varios formatos: (a) referencias cortas separadas del texto, a menudo comentadas y anotadas junto a una mesa; (b) citas breves incluidas en la narración; y (c) párrafos enteros de las notas de campo y transcripciones de las entrevistas. Las citas breves forman parte de las interpretaciones analíticas a lo largo de todo el estudio. Las citas más largas requieren que se guíe al lector «hacia» la cita y «hacia fuera» de ella para centrar su atención en la idea principal de la sección. Un ejemplo es el extracto 2. 8 (del capítulo 2), en el que cada interpretación se basa en citas seleccionadas de diferentes profesores. Las citas extensas, generalmente, se extraen del texto de la descripción y se hace referencia a ellas a través de alguna identificación del conjunto de los datos originales, como la fecha o el número de entrevistado.

Las estructuras de la narración y las representaciones visuales

Las estructuras de la narración y las representaciones visuales que forman parte de la práctica de la investigación, reflejan el proceso analítico. Entre las prácticas cualitativas existe la diversidad, especialmente en estudios críticos. Hablaremos de cada práctica haciendo hincapié en la investigación sobre los manuscritos.

Los **estudios etnográficos** proporcionan *una descripción, un análisis y una interpretación* de la cultura que el grupo comparte. El tramo más importante de la descripción holística se presenta por orden cronológico o por orden de la narración. El análisis puede centrarse en un día típico, un acontecimiento crítico, escenarios sociales seleccionados mediante el diálogo, o contar la «historia» desde diferentes perspectivas. El análisis puede también compararse y contrastarse con otros grupos dentro de la misma cultura. Los datos, transformados en modelos o temas, se resumen y proporcionan un retrato cultural, una síntesis de todos los aspectos de la vida del grupo ilustrando su complejidad (Walcott, 1994). La interpretación general y el significado se analizan a partir de los descubrimientos

relacionados con los resultados escolares o con los resultados reales (ver extracto 12.7). Las representaciones visuales de los descubrimientos pueden ser tablas, planos y diagramas.

Los **estudios fenomenológicos** de una experiencia vivida hacen hincapié en las descripciones textuales de *lo que ocurrió y cómo se experimentó el fenómeno*. Puesto que la experiencia la han vivido tanto el investigador como los entrevistados, los datos se recogen en el informe escrito a partir de la experiencia del investigador y de los entrevistados. El informe incluye una descripción de la experiencia de cada participante, incluyendo la del investigador, seguida de una *redacción descriptiva* y de la *esencia* de la experiencia. La esencia de la experiencia es una relación breve y altamente sintetizada del significado de la experiencia común. Las representaciones visuales son normalmente tablas de declaraciones literales de los entrevistados y una figura o tabla para la esencia de la experiencia.

EXTRACTO 12.7

PRESENTACIÓN EN FORMA DE DESCRIPCIÓN DE LOS DESCUBRIMIENTOS

En esta situación, las acciones de Mack y sus compañeros superan simples interpretaciones de desafío, inmadurez, resistencia y reproducción. Estos adolescentes tenían un conjunto de aspiraciones e intenciones que eran compatibles e incompatibles con las de sus parejas, familia y comunidad... Las vidas de estos jóvenes de raza negra... nos cuentan cómo se incorporan estas circunstancias críticas a las normas personales y públicas; todos comparten el mismo tipo de música, el mismo lenguaje, la misma manera de vestir e, incluso, hacen suyos negativos símbolos raciales/ étnicos, de clase e identificaciones de género para grupos como los negros (Kelley 1997b; Lusane 1993; Rose 1994).

Como conclusión, las voces e historias como las presentadas en este artículo cuestionan las opiniones peyorativas sobre la gente negra, de bajos ingresos, trabajadora y de clase media en los debates públicos sobre política...

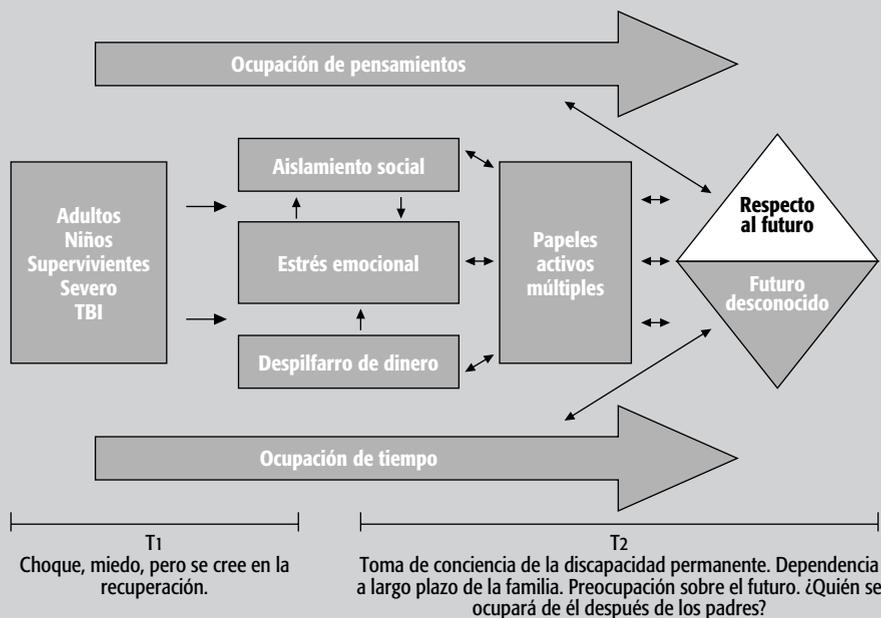
Fuente: Cousins, L. H. (1999), «Playing between classes: Americas' troubles with class, race, and gender in a black high school and community», *Anthropology & Education Quarterly*, 30 (3), 294-316.

La **investigación del estudio de caso**, similar a estudios etnográficos, contiene descripciones, análisis y generalizaciones naturalistas. Los estudios de caso, normalmente, utilizan un formato de informe con viñetas para relatar experiencias ajenas al lector (Stake, 1995). Los estudios de caso pueden estar a favor del análisis de la descripción y la interpretación de los párrafos entre un 60%–40%, o entre un 70%–30%. Hay descripciones extensas de los casos y sus contextos basados en una variedad amplia de fuentes de datos: documentación, archivo de grabaciones, entrevistas, observaciones de campo, observación del participante y objetos (Yin, 1994). Se presentan algunas consecuencias decisivas de manera que

el lector pueda apreciar la complejidad del caso. Dichas consecuencias se reco- gen a partir de la recopilación de casos entre los datos o el examen exhaustivo de un caso para detectar significados de consecuencias relevantes. Varias de las re- percusiones se analizan más profundamente con pruebas a favor o en contra. Finalmente, el investigador desarrolla una generalización naturalista (modelos) o «lecciones aprendidas» (Lincoln y Guba, 1985) que resultan útiles para los par- ticipantes y para los lectores que quieran aplicarlo en casos similares. Las repre- sentaciones visuales incluyen tablas, gráficos, figuras y diagramas integradores (ver extracto 12.8).

Los **estudios de teoría argumentada o fundamentada** subrayan una historia analítica a través de la descripción que se mantiene en segundo lugar en el *esque- ma teórico* o *concepto de la densidad* (elaboración). La historia analítica, también, especifica las variaciones de los fenómenos y las condiciones relevantes por múlti- ples causas. Las representaciones visuales son moldes condicionales, diagramas

EXTRACTO 12.8
ANÁLISIS DE LOS CONCEPTOS: REPRESENTACIÓN VISUAL
INCAPACIDAD DE DOMINIO



Fuente: Wood, W. M. (1993), *Long-term consequences of severe traumatic brain injury on adults: A qualitative study of families' perceptions of impact and public policy*. Discurso doctoral no publicado, Virginia Commonwealth University, Richmond.

integradores o modelos con contexto, condiciones causales, de entrevistas y consecuencias. Una representación visual acompaña las propuestas culminantes de la teoría argumentada. Las flechas indican el movimiento de los acontecimientos iniciales hacia las consecuencias tal y como vemos en el extracto 12.8. Gracias al énfasis de la discusión conceptual y al relato de la teoría basada en hechos para la bibliografía teórica, unos pocos lectores obtienen un juicio completo del estudio íntegro.

No obstante, algunos estudios de teorías basadas en hechos presentan una narración descriptiva más que una narración analítica que conecta las categorías y los avances de la teoría o densidad del concepto. Las propuestas teóricas pueden presentarse en forma narrativa o como una lista de declaraciones (ver extracto 12.9).

Los **estudios críticos** incluyen una investigación tan diversa como la investigación etnografía crítica, feminista y étnica, el análisis narrativo, la investigación de la acción participativa y la investigación sobre la acción. La mayor parte de las investigaciones etnográficas críticas y muchas de las investigaciones feministas y étnicas adoptan estructuras narrativas para realizar un informe a partir de investigación esencial. No obstante, los autores se preocupan por las distintas voces existentes, incluyendo la del autor en el discurso.

EXTRACTO 12.9

AFIRMACIONES SOBRE LA TEORÍA BASADA EN HECHOS

Este estudio fue una investigación cualitativa sobre el conocimiento y las opiniones, los papeles y los principios conductores de dos profesores ejemplares de ciencias en un instituto...

Los descubrimientos del estudio se resumen en las siguientes afirmaciones.

Afirmación 1: El conocimiento importante y la opinión de cada profesor quedan mejor representados como un grupo de principios de enseñanza.

Afirmación 2: Cada profesor ejerce múltiples papeles como enseñante, y cada papel está descrito por una metáfora diferente sobre el papel.

Afirmación 3: Los papeles como enseñante de cada profesor son consecuentes con su grupo de principios sobre la enseñanza.

Afirmación 4: Cada profesor tiene principios conductores que son constantes...

Fuente: Lloyd, J. M. (1999), *Knowledge and beliefs, roles, and guiding principles of two exemplary high school science teachers and model for teacher reflection*. Discurso doctoral no publicado, Virginia Commonwealth University, Richmond.

L. Richardson (1990, 1998) recuerda a los autores que el pensamiento posmoderno enlaza el lenguaje, la subjetividad, la organización social y el poder. El lenguaje no «refleja» la realidad social, sino que produce significados y por lo tanto, crea una realidad social. La transformación del texto de campo en texto narrativo es el lugar en el que se construyen el sentido del *sí mismo* y de la

subjetividad personal. Al destacar la escritura reflexiva, el autor es libre de escribir como un único texto en el que todo se dice para todo el mundo una vez.

Un número de científicos sociales cualitativos están escribiendo formularios experimentales de representaciones sugerentes⁶. Algunos de estos escritos son narraciones del sí mismo, representaciones etnográficas novelescas, representaciones poéticas, dramas etnográficos y géneros mezclados. Numerosas revistas cualitativas escolares han publicado estudios que empleaban estructuras narrativas experimentales. Muchos investigadores cualitativos han sido influidos por los formatos experimentales y han escrito de manera más reflexiva y centrada en sí mismos; también redactan sus propios textos más profunda y complejamente mientras que sus propias voces crecen.

RESUMEN

Los siguientes puntos resumen el proceso y las técnicas del análisis de datos cualitativos:

1. El análisis inductivo implica que las categorías y los modelos surgen a partir de los datos más que venir impuestos por los datos principales de la recopilación de datos.
2. El análisis de datos implica fases cíclicas, análisis de descubrimiento en el campo, identificación de los temas que se convierten en categorías y síntesis de los modelos entre las categorías.
3. Los estilos analíticos del análisis de datos cualitativos incluyen técnicas predefinidas, estilos plantilla, de edición y cristalización.
4. El análisis del descubrimiento de las estrategias, realizado durante la recopilación de datos, incluye escribir muchos «comentarios del observador» y resúmenes, jugar con las ideas, examinar la bibliografía y emplear metáforas y analogías.
5. El análisis interno ayuda en la toma de decisiones sobre la recopilación de datos y en la identificación de los temas y significados que se repiten.
6. La codificación es el proceso que consiste en la división de los datos en partes a través de un sistema de clasificación –segmentar los datos en temas o usar categorías predeterminadas para formar subcategorías más pequeñas–.
7. La técnica analítica es una comparación constante (comparar y contrastar) para determinar las características distintivas de cada tema y categoría.
8. Las categorías predeterminadas derivan del problema de investigación, una guía de la entrevista, bibliografía y experiencia personal del investigador.
9. Un segmento de datos (una palabra, frase, párrafo o página) es comprensible por sí sólo y contiene una idea, episodio o información relevante para el estudio.

⁶ Ver Richardson, 1998, p. 355-361, para ejemplos citados y revistas.

10. Un tema es un nombre descriptivo para la materia del tema, no para el significado del segmento.
11. Los investigadores perfeccionan su sistema de clasificación de temas provisional a lo largo del estudio.
12. Una categoría es un concepto abstracto para significar temas similares.
13. Las técnicas para desarrollar categorías incluyen las preguntas analíticas, la comparación mental de situaciones similares y diferentes, y el análisis de las «banderas rojas».
14. Las categorías emic representan el punto de vista de los participantes o interiores; las categorías etic representan los puntos de vista de la ciencia social o exteriores.
15. Un modelo es una relación entre categorías.
16. Las técnicas para la búsqueda de modelos incluyen la triangulación, ordenar y clasificar las categorías, analizar las evidencias discrepantes o negativas, construir representaciones visuales y realizar análisis-cruzados lógicos.
17. Los modelos son explicaciones posibles cuando cuentan con datos como prueba y los modelos alternativos no son razonables.
18. El manejo de datos, realizado manualmente o a través de la asistencia por ordenador, es un sistema para recuperar conjuntos de datos y para unir los datos codificados en archivadores o expedientes informáticos.
19. Los estudios cualitativos adoptan diferentes formas según su auditorio, reconocen la presencia del autor y emplean estructuras narrativas relacionadas con la tradición de investigación escogida.
20. Los estudios cualitativos presentan el contexto y las citas del lenguaje de los participantes como datos.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. El análisis inductivo es...
 - a. imponer categorías en los datos principales para la recopilación de datos.
 - b. probar hipótesis.
 - c. presentar datos en tablas estadísticas.
 - d. desarrollar categorías ante todo a partir de los datos.
2. El análisis de datos en la investigación cualitativa...
 - a. se realiza tras haber recopilado todos los datos.

9. Los resultados cualitativos se presentan a través de...
 - a. la omisión del contexto del fenómeno estudiado.
 - b. informes analíticos que contienen el lenguaje del investigador.
 - c. citas de las palabras y afirmaciones de los participantes.
 - d. parafrasear el lenguaje de los participantes para cambiar el significado.
10. Una presentación *inexacta* que concierne a las estructuras narrativas es...
 - a. estudios cualitativos codificados para diferentes auditorios.
 - b. la presencia del autor se deja ver a través de varias referencias intelectuales.
 - c. la tradición de la investigación que subraya una narración completamente descriptiva es una teoría argumentada.
 - d. Algunos estudios críticos adoptan una estructura de informe fundamental y otros emplean formas experimentales de representaciones sugestivas

Problemas de aplicación

1. A continuación aparecen comentarios de opiniones de profesores sobre la preparación y las prácticas del jardín de infancia (M. L. Smith y L. S. Shepard, 1988, p. 316-319). Identifique los números de las líneas para las categorías de:
 - opiniones sobre el desarrollo de los niños;
 - opiniones sobre la posibilidad de «ponerse al día»;
 - opiniones sobre las posibilidades de influir en una preparación de los niños para la escuela;
 - opiniones sobre qué pueden hacer los profesores.

Elabore una lista de cualquier otra categoría que observe en los datos.

Sr. Willis:

- 1 «Puesto que el desarrollo constituye una manifestación fisiológica, las proporciones de
- 2 desarrollo son uniformes, continuas, sin esfuerzos ni discontinuidad.
- 3 El niño que lleva un retraso de 6 meses en septiembre llevará el mismo retraso en
- 4 junio... Hay pocas probabilidades de que un niño que tiene un desarrollo mental inferior
- 5 que sus compañeros, salve esta diferencia que les separa... La intervención es
- 6 inútil con un niño sin suficiente desarrollo mental. La ayuda extra y los remedios

- 7 causan presión, frustración y compensación. Los profesores no pueden influir
- 8 en las destrezas psicomotoras, la capacidad de atender, la madurez social,
- 9 etc... Los profesores pueden proporcionar al niño más tiempo para madurar; colocar al
- 10 niño en un jardín de infancia de desarrollo mental o preescolar, enviarlo a casa durante otro
- 11 año; colocar al niño en un grupo lento de clase; reducir la instrucción para bajar
- 12 el nivel de frustración, tener expectativas más bajas, aumentar la autoestima, usar
- 13 manipulaciones; retenerlo en el jardín de infancia; proporcionarle asistencia académica es
- 14 irrelevante y perjudicial».

Srta. Johnson:

- 15 «Dentro de los amplios límites de la edad cronológica, la rapidez de los niños es una
 - 16 función de sus experiencias, aprendizaje del programa y el entorno... Un niño
 - 17 que está menos preparado que sus compañeros puede salvar esta diferencia si se dan las
 - 18 circunstancias educativas necesarias; la asistencia académica requerida... El profesor
 - 19 puede hacer una diferencia igual que el padre y cualquier otro aspecto del entorno;
 - 20 dentro de un grupo de capacidades del alumno, lo que el alumno aprende es ampliamente una
 - 21 función de oportunidades y experiencias... El profesor puede proporcionar
 - 22 ayuda académica adicional; adaptar las diferencias en la puntuación; mantener las
 - 23 expectativas altas, reforzar y enseñar; trabajar duro y animar al alumno a trabajar
 - 24 duro».
2. Ahora, retroceda y observe los significados de cada categoría. ¿Hay opiniones que están relacionadas con otras opiniones, es decir, hay categoría(s) relacionada(s) con otra(s) categoría(s) para sugerir un modelo? Formule cualquier modelo que descubra en los datos.

IV

Investigación analítica

Frecuentemente, los educadores preguntan: ¿por qué es importante la historia de la escolarización y de las prácticas educativas? ¿Qué significa un concepto o un término? ¿Cómo puede el conocimiento del pasado iluminarnos y darnos información en debates públicos sobre educación y en procesos de toma de decisiones?

La investigación cualitativa quedó clasificada en el capítulo 2 como *investigación interactiva* o *investigación no interactiva* denominada *investigación analítica*. Hablaremos de la investigación cualitativa que es tradicionalmente no interactiva en la parte IV.

La investigación analítica incluye el análisis de conceptos y de acontecimientos históricos y los procesos de elaboración y desarrollo de políticas educativas. Las fuentes de investigación histórica son documentos, testimonios orales y vestigios del pasado. El investigador utiliza técnicas especializadas para investigar y localizar documentos en archivos, manuscritos, bibliotecas y colecciones privadas.

Hablaremos tanto de las características generales de la investigación analítica como de los procedimientos específicos empleados a través del análisis de conceptos y del análisis histórico. También presentaremos los patrones de credibilidad para esta tradición de investigación e ilustraremos el valor de la búsqueda histórica.

Análisis de conceptos e investigación histórica



PALABRAS CLAVE

historiografía

historia oral

biografía

biografía interpretativa

testimonios orales

documentos

restos materiales

hechos

generalizaciones

interpretación analítica

análisis conceptual

fuente primaria

fuente secundaria

crítica externa

crítica interna

Una manera de comprender las prácticas educativas actuales es conocer cómo han evolucionado dichas prácticas y esclarecer los asuntos relacionados con esta evolución. ¿Cuántas veces los educadores y no educadores han realizado declaraciones o justificado decisiones basándose en lo que consideran que ha sucedido en el pasado? Las explicaciones de las ideas o de los conceptos, de los acontecimientos y de las políticas educativas del pasado sugieren nuevas percepciones de las tendencias y de los acontecimientos educativos actuales y nuevas consecuencias educativas (Strickler, 1992; Mason y otros, 1997).

La investigación analítica, como una forma de indagación cualitativa, proviene fundamentalmente de las disciplinas de la filosofía (el significado de los conceptos), la historia y la ciencia política. El análisis de conceptos y la investigación histórica son tradicionalmente investigaciones de documentos no interactivas. No obstante, algunas formas de investigación histórica, como la historia oral y la biografía interpretativa, emplean técnicas interactivas complementadas con documentos y registros. La investigación analítica describe e interpreta el pasado remoto o reciente a partir de las fuentes seleccionadas.

Tanto las formas interactivas como no interactivas de investigación cualitativa comparten las generalizaciones limitadas al contexto, a la orientación del descubrimiento, al diseño del estudio de caso emergente, al énfasis holístico (las cualidades de las partes unifican el fenómeno entero), y a la no interferencia en el escenario natural y el análisis inductivo de datos¹.

¹ Ver Sherman y Webb, 1988, para una comparación de las tradiciones de investigación cualitativa como la etnografía, la teoría fundamentada, la filosofía, la historia y otros.

CARACTERÍSTICAS GENERALES DE LA INVESTIGACIÓN ANALÍTICA

Subrayar las variaciones de la investigación analítica es una característica metodológica común que distingue los estudios analíticos de otro tipo de investigaciones educativas. Dichas características metodológicas incluyen un tema de investigación relacionado con acontecimientos pasados, fuentes primarias como datos, técnicas de crítica empleadas en la investigación de hechos y explicaciones interpretativas. Puesto que dichas características son generales, pueden aplicarse de diferentes maneras en un estudio determinado.

La **historiografía** es el estudio de los procedimientos que los distintos historiadores manejan en su investigación, y de las revisiones y variables interpretaciones del pasado. Nos limitaremos a hablar sobre las características generales y procedimientos más específicos descritos más adelante en este capítulo con ejemplos diseñados a partir de acontecimientos políticos e históricos educativos de Estados Unidos. Muchos de estos estudios requieren el análisis de conceptos educativos, es decir, el significado del lenguaje utilizado.

Temas de análisis

Los temas históricos. Los temas históricos incluyen un amplio grupo de temas nuevos y recurrentes de interés. Las siguientes cuestiones ilustran la diversidad de la investigación histórica:

1. Movimientos: educación progresiva, aprendizaje a lo largo del ciclo vital.
2. Instituciones: educación pública, jardín de infancia, guardería.
3. Conceptos: escolaridad, el niño, alfabetización, profesionalidad.
4. Biografías de educadores influyentes: John Dewey, Phillis Wheatley.
5. Historia comparativa de educación internacional: comparación del sistema educativo estadounidense con los de otras naciones.
6. Formas alternativas de escolaridad: instrucción en casa, educación a distancia.
7. Componentes de la educación: asuntos financieros, personal, acreditación, currículum, contratos, organización, administración, métodos de instrucción y materiales.
8. Educación cultural y de minorías: género, ética, minoría, educación bilingüe.
9. Regionalismo en la educación estadounidense: regiones geográficas; sistemas educativos estatales; educación urbana/rural.
10. Otros temas: recopilación (restauración) de documentos con anotaciones; narración cronológica de series de acontecimientos.

Más recientemente, los historiadores se han centrado en los aspectos educativos de: (a) la familia, la iglesia y las asociaciones profesionales; (b) el medio urbano y el medio rural; (c) las instituciones especiales como los reformatorios o centros de menores, los orfanatos y los centros juveniles; y (d) la cultura popular como la televisión, las canciones y la literatura (Cohen, 1976).

Una tendencia que continúa actualmente es la aplicación de conceptos y métodos tradicionales de otras disciplinas, como la sociología, la economía, la psicología y la antropología a la investigación histórica (Kaestle, 1992). El pasado histórico puede llegar a ser un tiempo tan reciente como el año anterior o tan lejano como hace siglos de los que el historiador no cuenta con una experiencia personal. Las monografías y los periódicos especializados reproducen investigaciones históricas.

Historia oral. Una forma de investigación histórica es la **historia oral** que registra las palabras y los testimonios de los individuos. Un estudio puede centrarse en la grabación de relatos populares e historias de una región o un grupo cultural. Las entrevistas orales de personas que han sido testigos o han participado en acontecimientos históricos importantes son grabadas en casete y las transcripciones resultantes proporcionan un registro escrito. Los historiadores orales recopilan y preservan la historia oral antes de que se «pierda» para las generaciones futuras².

Biografía. Un estudio que se centra en un *individuo*, especialmente en los momentos más importantes de su vida, tal como se narra al investigador o se registra en documentos o materiales archivados, es una **biografía**. La biografía, como un tipo de investigación, puede incluir biografías individuales, autobiografías, relatos e historias orales de la vida de algunas personas (Smith, 1998). Estos documentos estudian «vidas menores, grandes vidas, vidas frustradas, vidas cortas o vidas milagrosas con sus logros desconocidos» (Creswell, 1998, p. 48).

Un investigador puede acercarse a su tema como si se tratara de una biografía *clásica* o de una biografía *interpretativa*. Una biografía clásica es aquella en la que al investigador le preocupa la validez y crítica de las fuentes primarias y el desarrollo de una base «válida» para las explicaciones³. Una **biografía comentada** es un estudio en el que la presencia del investigador queda patente en la narración así como su punto de vista. Los biógrafos interpretativos reconocen que, en cierto modo, el escritor «crea» a la persona de la narración del mismo modo que los individuos se crean a sí mismos al contar o registrar sus historias vitales (Denzin, 1989). En ambos tipos de biografía, el investigador debe tener conocimiento del contexto histórico del individuo

² Véase la publicación anual de la *Oral History Review*.

³ Véase Angroisino (1989) y Lomask (1986) para un enfoque más tradicional de la investigación biográfica.

para, así, poder interpretar la vida de la persona dentro de tendencias sociales o culturales más amplias.

Temas políticos. Los temas políticos incluyen dos áreas distintas:

1. El análisis político sobre la naturaleza y el empleo del poder y la influencia en la administración educativa.
2. El contenido político y el impacto en educación de acciones gubernamentales específicas (Mitchell, 1988)⁴.

La primera de estas áreas temáticas tiene como objetivo explicar la política escolar a través de la observación de la distribución de poder entre varios implicados en el sistema, y a través del seguimiento de las interacciones entre estos grupos de influencia para revelar los procesos de toma de decisiones. Esta investigación, denominada «política educativa», se centra en el poder de los grupos de interés profesional, redes sociales informales que promueven iniciativas políticas, el pequeño número de personas clave que da forma a las promulgaciones legislativas y las variaciones de valores culturales entre los niveles de la administración educativa.

Los estudios de contenido político se concentran en políticas o asuntos específicos más que en las relaciones de poder. No obstante, en un nivel teórico, los estudios políticos son muy diversos. La investigación sobre el contenido político puede centrarse en cualquiera de los mecanismos de control para mejorar las escuelas, como la organización escolar y su forma de gobierno, la financiación de las escuelas, los exámenes y la evaluación de los alumnos, la definición del programa escolar, la formación y la acreditación del personal, el desarrollo y selección de los materiales curriculares y los edificios y las instalaciones escolares. La investigación política también se centra en cómo el contenido de las políticas cambia en las distintas fases de su elaboración, como la articulación de un proyecto, la incorporación de los grupos interesados y la formación de coaliciones, el reparto de poder y de recursos para decretar políticas, la transformación de las leyes o reglas en reglamentos, la implantación de la política en la práctica y la evaluación del grado en que las políticas han sido implantadas según el proyecto inicial y/o han dado los resultados esperados. Muchos estudios indican que el contenido de una política cambia de una fase a otra.

Gracias a los estudios analíticos, conocemos y comprendemos la historia educativa pasada y los acontecimientos políticos. Las ideas y conceptos principales quedan aclarados por su significado. Las preguntas de investigación se centran en acontecimientos (quién, qué, cuándo, dónde), cómo tuvo lugar un acontecimiento (descriptivo) y por qué ocurrió (interpretativo).

⁴ El análisis político, un tipo de investigación diferente, se expone en el capítulo 14.

Tipos de fuentes

Los datos de estos estudios son fuentes escritas, muchas de las cuales se han conservado en archivos, en manuscritos, en recopilaciones personales o en bibliotecas. Las fuentes son documentos, testimonios orales y restos materiales. Todas estas fuentes se clasifican generalmente como documentos. Un estudio puede requerir uno o varios tipos de fuentes.

1. Los **documentos** son *registros de acontecimientos pasados*. Son materiales escritos o impresos que pueden ser oficiales y no oficiales, públicos o privados, publicados o no publicados, realizados intencionadamente para preservar una recopilación histórica o para conseguir un objetivo práctico inmediato. Los documentos pueden ser cartas, diarios, testamentos, recibos, mapas, revistas, periódicos, actas oficiales, edictos, reglamentos o leyes.

Un tipo especial de documento son los *registros cuantitativos*, que pueden incluir registros de inscripción o matrícula, registros de contratación del personal, listas de miembros, documentos del censo, listados fiscales, registros electorales, presupuestos, datos sobre resultados de tests y cualquier recopilación de datos numéricos. El compendio de los datos, cuando son auténticos, hace que sea más fácil describir y analizar la información. La dificultad de utilizar los registros cuantitativos aumenta cuanto más remoto es el periodo estudiado. Como escribió Aydelotte (1986, p. 806), «las presentaciones estadísticas formales son viables sólo para un grupo limitado de problemas históricos»; sin embargo, algunos datos políticos, económicos y demográficos se han tratado cuantitativamente con éxito⁵.

2. Los **testimonios orales** son *registros del lenguaje oral*. Los testimonios orales de las personas que han presenciado acontecimientos de importancia educativa se graban en casete y se realizan e identifican las transcripciones literales. Los testimonios orales son entrevistas autobiográficas o en profundidad y, o bien constituyen la principal fuente de evidencia o bien se usan para complementar las pruebas documentales. Dichos testimonios pueden ser recogidos por los participantes o testigos de la fundación de una nueva institución, la aprobación de una ley educativa o la implantación de una política.

3. Los **restos materiales** son cualquier *objeto* que proporcione información sobre el pasado. Aunque puede que no proporcionen información directa sobre el pasado, sus propiedades visuales y físicas pueden facilitar pruebas. Los restos pueden ser libros de texto, edificios, equipamiento, gráficos, exámenes, las pruebas físicas presentadas ante un tribunal o los objetos físicos en la elaboración de políticas. La tabla 13.1 ilustra los tipos de fuentes para la investigación histórica y de elaboración de políticas.

⁵ Un ejemplo se encuentra en J. Dpughtery (1998). El procedimiento está descrito en J. Audelotte (1989) y Burton. O. V. y Finnegan, R. (1990).

TABLA 13.1: TIPOS ILUSTRATIVOS DE FUENTES PARA LA INVESTIGACIÓN ANALÍTICA

Fuente	Investigación histórica	Investigación sobre la elaboración de políticas
Documentos	Cartas Diarios Facturas y recibos Autobiografías Periódicos Revistas Boletines Catálogos Películas Grabaciones Registros personales Registros institucionales Presupuestos Registros de matriculación Registros de graduación	Registros gubernamentales oficiales y no oficiales federales, estatales, locales y escolares <ul style="list-style-type: none"> • Vistas legislativas, debates, informes, publicaciones • Informes, actas y circulares de comités y comisiones • Informes de agencias • Datos estadísticos, presupuestos • Reglamentos, directivas • Leyes • Registros electorales Discursos de miembros del gobierno Decretos y edictos
Testimonios orales	Participantes en un acontecimiento histórico Familiares de una persona fallecida	Participantes en los procesos de elaboración de políticas: <ul style="list-style-type: none"> • administradores • consejo escolar • miembros • gobierno • directores • personal profesional • grupos especiales de interés • asociaciones profesionales
Restos materiales	Libros de texto Edificios Mapas Equipamiento Muestras del trabajo de los alumnos Muebles Materiales educativos	Objetos físicos presentados en vistas legislativas o en otros procesos de la elaboración de políticas: <ul style="list-style-type: none"> • gráficos • diagramas • restos materiales históricos

La búsqueda de pruebas

La búsqueda de pruebas basadas en datos empieza con la localización de las fuentes. Mientras que los investigadores cuantitativos crean los datos al asignar

instrumentos a una muestra, el historiador depende de las fuentes que se han conservado. Dichas fuentes pueden o no haber sido catalogadas e identificadas para acceder a ellas fácilmente. Encontrar y valorar las fuentes históricas es un ejercicio detectivesco. Implica «lógica, intuición, persistencia y sentido común, la misma lógica, intuición, persistencia y sentido común que se utilizarían para conseguir datos contemporáneos...» (Tuckman, 1998, p. 258).

La búsqueda de pruebas requiere localizar las fuentes primarias y secundarias. Las fuentes primarias son documentos o testimonios de testigos oculares de un acontecimiento. Las fuentes secundarias son documentos o testimonios de individuos que no observaron o participaron realmente en el acontecimiento y que, por lo tanto, hablan sobre lo que han oído decir. Los testimonios presenciales se valoran más que las fuentes secundarias, pero ambos tipos de fuentes deben someterse a técnicas de crítica.

Las técnicas de crítica valoran la autenticidad y la veracidad de la fuente. La autenticidad determina si la fuente es un documento genuino, falsificado o una variante del documento original. La veracidad de la fuente hace referencia a la exactitud de las declaraciones encontradas en la fuente. El investigador localiza las fuentes más auténticas y verosímiles para indagar los hechos. Los **hechos** son las piezas más fieles entre los diferentes tipos de fuentes auténticas. Las interpretaciones de un acontecimiento se basan en los hechos.

Generalizaciones analíticas e interpretaciones

Los estudios analíticos sugieren **generalizaciones** que son resúmenes de datos basados en pruebas. Cuando aplicamos la lógica inductiva a una serie de generalizaciones sobre un acontecimiento, surge una explicación. Una **interpretación analítica** es una explicación causal de un acontecimiento específico. Términos como *causa, porque, ya que, debido a y por la razón de que* connotan explicaciones. Las explicaciones analíticas implican causas con afirmaciones como «bajo estas circunstancias, no es sorprendente que...», «naturalmente, llegados a este punto, él...» y «era inevitable que...». Términos como *imposibilidad, necesidad, influencia, impulso, desarrollo, consecuencia* y *motivos* y los verbos como *tender a, producir, provocar, estimular, forzar* y *obedecer*, aluden a algún tipo de explicación.

Las interpretaciones analíticas válidas difieren de las explicaciones de sentido común de las siguientes formas:

1. **Una generalización analítica compila los hechos separados que afirman que un acontecimiento tuvo lugar.** El investigador se interesa por las circunstancias particulares bajo las que aconteció un hecho y reconoce que es posible que dichas circunstancias no vuelvan a repetirse nunca. Las explicaciones analíticas se diferencian de las explicaciones de sentido común

porque aquéllas requieren juicios críticos mediante el empleo del conocimiento especializado sobre la época en la que el acontecimiento tuvo lugar y sus circunstancias.

2. **Las interpretaciones sugieren múltiples causas para un acontecimiento singular.** Las interpretaciones de las conexiones entre las generalizaciones son normalmente causales. La causa y el efecto no se tratan de modo absoluto y, frecuentemente, dependen del orden cronológico de las incidencias que culminan en un hecho concreto. Cuanto más complejos sean los acontecimientos y más amplio sea el intervalo de espacio y tiempo, más necesaria será la interpretación del historiador. Normalmente, un acontecimiento de cualquier magnitud y complejidad necesita muchas generalizaciones para ser interpretado.
3. **Cuando las interpretaciones analíticas está justificadas o apoyadas por hechos expuestos en el estudio, las explicaciones son consideradas como válidas.** Los historiadores dirán: «si no confías en mi explicación, estudia los hechos». La especificación de los detalles –quién, qué, cuándo, dónde y cómo– es precisamente la característica que la diferencia de explicaciones estadísticas (Gardiner, 1952). No obstante, los historiadores, raras veces, afirman que conocen todas las circunstancias. En su lugar, un estudio contiene un conjunto de hechos e ideas asociados que, cuando se presentan en un espacio determinado, no deja preguntas sin respuestas *dentro* del análisis. Sin embargo, se pueden formular muchas preguntas sobre el tema desde *fuera* del análisis en otros estudios (Barzun y Graff, 1992). La figura 13.1 describe el proceso de investigación analítica.

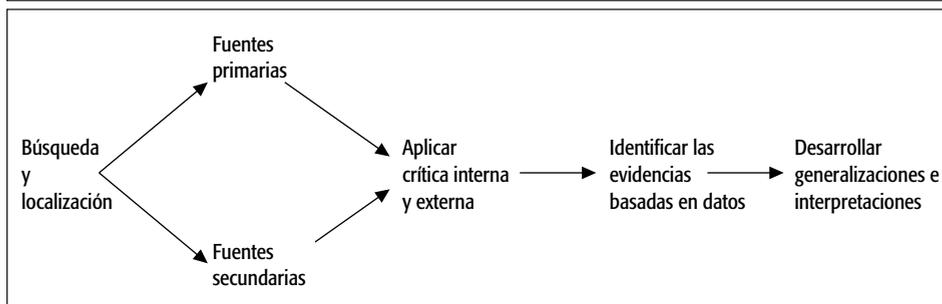
Aproximaciones a la investigación analítica

La investigación *histórica tradicional* se ha centrado en la investigación de las «causas» de los acontecimientos del pasado (Carr, 1967). La mayoría de los historiadores defienden que los acontecimientos históricos son únicos e irrepetibles. En la identificación de las «causas» de los acontecimientos, los historiadores pueden basarse en la cronología de una serie de hechos, asumir la historia conocida por todos, hacer referencia a las ideologías convincentes, citar avances tecnológicos y científicos o centrarse en influencias sociológicas, psicológicas, políticas o económicas.

Tal y como escribió el conocido historiador Gottschalk (1969), el «pasado completo» puede ser conocido por el investigador sólo «a través de los registros que se conservan... y la mayor parte de la historia como registro es sólo la parte que se conserva de la parte registrada, de la parte recordada, de la parte observada de la totalidad» (p. 45-46). Parece que no existen causas absolutas esperando ser descubiertas y ofrecidas como explicación por el investigador. En lugar de

ello, los historiadores escriben en diversos niveles de abstracción, a diferentes distancias cronológicas del acontecimiento pasado, con distintos objetivos, en diferentes contextos y desde puntos de vista distintos. No resulta sorprendente que diferentes estudios propongan explicaciones del pasado aparentemente contradictorias pero válidas.

FIGURA 13.1: EL PROCESO DE LA INVESTIGACIÓN ANALÍTICA



La historia es, a menudo, reinterpretada de acuerdo con la importancia que adquieren los intereses o preocupaciones de los investigadores y del público en general, o conforme se dispone de nuevos documentos. Así pues, muchos historiadores reconocen la falibilidad de sus interpretaciones, incluso si son razonables, en un momento dado. Como historiadores *interpretativos*, son conscientes de sus propios valores, opiniones e intereses sobre un tema. Frecuentemente, son conscientes de ello cuando leen cómo otros investigadores han estudiado el tema.

Durante las últimas décadas, algunos historiadores, llamados historiadores *re-visionistas* o *reconstruccionistas* han cuestionado los puntos de vista convencionales o populares del pasado. Sugieren que las escuelas son un instrumento de control político y estratificación social que promueven directa o indirectamente la discriminación contra la mujer y los grupos étnicos. Algunos historiadores revisionistas se parecen a los investigadores críticos de los que hablamos en la parte III.

Las siguientes secciones de este capítulo describen la metodología empleada en la investigación analítica. Describiremos las técnicas del análisis conceptual y los procedimientos para estudiar los acontecimientos históricos y políticos pasados.

ANÁLISIS DE CONCEPTOS EDUCATIVOS

Un **análisis conceptual** es un estudio que esclarece el significado de un concepto a través de la descripción de los significados esenciales o genéricos, sus

distintos significados y el uso adecuado a ese concepto. Al presentar un análisis del concepto, el estudio nos ayuda a comprender la manera de pensar de la gente sobre la educación. El foco de interés se encuentra en el significado del concepto, no en los valores personales del investigador o en la información basada en datos. Por ejemplo, la pregunta: «¿qué objetivos tiene la educación?», no se contesta a través de la recopilación de datos que muestren que los educadores poseen o no objetivos, ni haciendo una declaración de valores de que los educadores deberían tener objetivos. En su lugar, el analista empieza preguntándose: «¿el hecho de tener un objetivo constituye una parte necesaria y fundamental de nuestro concepto de ser un educador?» o «¿puede una persona ser educadora sin tener un objetivo?» (Soltis, 1978, p.15). Plantearse dichas preguntas permite al analista llegar al significado de la educación, examinando cuidadosamente algunas ideas relacionadas con el concepto de los objetivos como un aspecto de la educación. El analista asume una posición neutral mientras que analiza un concepto antes de adoptar un enfoque o de recopilar información basada en datos.

Podemos emplear tres estrategias para analizar conceptos como *educación*, *alfabetización*, *conocimiento*, *enseñanza*, *aprendizaje*, *igualdad de oportunidades* y *corrección formal*. Soltis (1978) nos muestra análisis genéricos, diferenciales y de condiciones. Cada una de estas estrategias empieza con una pregunta principal diferente y utiliza técnicas específicas para explicar el concepto.

1. ***Un análisis genérico identifica el significado esencial de un concepto.*** El análisis aísla los elementos que distinguen al concepto de los otros términos. Para aclarar el concepto de *disciplina académica*, se debe, por tanto, comparar la historia, las matemáticas y la física como evidentes ejemplos típicos con la economía doméstica, la agricultura intensiva y el esquí acuático como contraejemplos para llegar al significado genérico de *disciplina académica*.
2. ***Un análisis diferencial distingue entre los significados básicos del concepto y proporciona una idea más clara del dominio lógico que abarca el concepto.*** El análisis diferencial se usa cuando un concepto parece tener más de un significado estándar y cuando la base para diferenciar los significados no resulta evidente. La pregunta principal es: «¿cuáles son los significados básicos (diferentes) del concepto?». Uno se podría preguntar, por ejemplo, «¿cuáles son los diferentes significados de la materia?» y analizar el concepto de materia clasificando intuitivamente los usos frecuentes con ejemplos concretos, como *Silas Marner*, sistema solar, asignaturas escolares, conocimiento y habilidades. Las características que diferencian cada tipo de materia escolar sirven para separar claramente los tipos y se puede desarrollar una tipología. Un ejemplo de análisis diferencial es la distinción entre investigación básica, aplicada y evaluativa expresada en el capítulo 1, que pretende aumentar la comprensión del concepto de *investigación*.

3. *Un análisis de condiciones identifica las condiciones necesarias para el uso apropiado del concepto.* La pregunta principal es: «¿bajo qué circunstancias contextuales sería cierto que el concepto está presente?». El análisis de las circunstancias empieza proporcionando un ejemplo que cumple las condiciones necesarias del concepto pero que puede perder su carácter de ejemplo si modificamos el contexto. Esto obliga a la revisión o al rechazo de la condición y conduce a condiciones adicionales con otros ejemplos o contraejemplos. El objetivo del análisis de condiciones «es producir un conjunto de condiciones necesarias y suficientes para la aplicación correcta de un concepto en cualquiera de sus muchos contextos posibles» (Soltis, 1978, p. 65).

En el análisis de conceptos educativos resulta crucial la selección de empleos representativos del concepto y de contraejemplos. El analista utiliza un muestreo intencional al escoger ejemplos que demuestran los significados implícitos en el lenguaje, que se analizan lógicamente con posterioridad. Los ejemplos pueden establecerse a partir de usos de conceptos comúnmente aceptados. Puesto que se utilizan frecuentemente grupos de ejemplos, el análisis de conceptos educativos puede desembocar en una repetición del análisis y en claridad conceptual posterior.

El análisis de conceptos educativos se aplica al estudio de conceptos educativos y a la investigación histórica, legislativa y política. Un estudio de la evolución de la escuela pública, el oportuno procedimiento educativo o los grupos que influyen en la implantación de una ley, debe determinar primero el significado de escuela pública, el propio procedimiento y los grupos de influencia.

ANÁLISIS DE LA HISTORIA EDUCATIVA Y DE LOS ACONTECIMIENTOS POLÍTICOS

La investigación analítica requiere una aplicación sistemática de los procedimientos metodológicos para comentar un problema histórico, localizar y criticar las fuentes e interpretar los hechos para exponer las explicaciones causales. Para llevar a cabo una investigación histórica y política es necesario tener una formación especializada. El historiador actúa de manera circular debido a las interrelaciones del problema de investigación, las fuentes, las críticas, el análisis y las explicaciones.

El tema y el enunciado del problema

El investigador empieza con un tema inicial, como por ejemplo, un periodo histórico, una persona, una idea, una práctica, una institución o una determinada

política. Conforme el analista obtiene conocimientos de los antecedentes, el tema se va definiendo de manera más exacta. Simultáneamente, el analista anota posibles fuentes primarias relevantes para el tema. El enunciado del asunto delimita y centra el estudio de investigación. El asunto debe ser lo suficientemente concreto como para examinar el acontecimiento en detalle, pero lo suficientemente amplio como para identificar los patrones para su interpretación.

Para expresar el tema de la investigación, el historiador inicialmente revisa las fuentes secundarias, de forma amplia y detenida, para obtener un conocimiento de los antecedentes. Este conocimiento le da una visión aproximada de la extensión del tema, la investigación previa sobre el problema, las lagunas de conocimiento y las posibles fuentes. Este conocimiento se obtiene a partir de los libros de texto, monografías, enciclopedias y otros trabajos de referencia, tesis y revistas especializados. Las bibliografías generales citan fuentes secundarias. Algunas bibliografías especialmente útiles para el historiador son *A Guide to Historical Literature* (Una guía para la literatura histórica), *The Historian's Handbook: a Descriptive Guide to Reference Works* (El manual del historiador: una guía descriptiva para trabajos de referencia) y *A Bibliography of American Educational History* (Bibliografía de la historia educativa estadounidense). Las fuentes secundarias para la investigación política se localizan a través de bibliografías históricas generales, y algunos ejemplos podrían ser *The Study of Politics and Education: A Bibliographic Guide to Research Literature* (El estudio de la política y la educación: una guía bibliográfica para la literatura de investigación).

Limitar y formular un tema es un esfuerzo continuo. El enunciado del problema se expresa más breve y claramente al final de la investigación, cuando las fuentes han sido recopiladas, analizadas e interpretadas. Aspectos a tener en cuenta en la delimitación de un tema son la disponibilidad y la accesibilidad de las fuentes primarias; los intereses del analista, el conocimiento especializado y el tiempo para completar el estudio, y el tipo de análisis que se va a realizar.

El enunciado del problema histórico indica el acontecimiento, persona, institución o política concretos. Los problemas quedan delimitados por el periodo de tiempo, la localización geográfica y el punto de vista del análisis (ver el extracto 13.1).

Localización y crítica de las fuentes

Las fuentes primarias son esenciales para la investigación analítica, pero las secundarias se emplean selectivamente cuando son necesarias. Tanto las primarias como las secundarias están sujetas a técnicas de crítica. Las fuentes para un estudio se citan en las referencias bibliográficas y, frecuentemente, en las notas a pie de página de un estudio. La crítica de las fuentes puede aparecer en el texto de un estudio o en un apéndice metodológico.

Clasificación de las fuentes. Una **fente primaria** es un testimonio escrito u oral de un testigo presencial o de un participante, o un registro realizado por algún medio mecánico presente en el acontecimiento, como una grabadora, una transcripción o una fotografía. Las fuentes primarias para una biografía histórica son los documentos personales y públicos de la persona y los recuerdos materiales de su vida. Las fuentes primarias para la investigación sobre la elaboración de políticas son los documentos de acciones gubernamentales y testimonios orales de testigos presenciales. Una fuente primaria es original en el sentido de que contiene los datos de primera mano de los testigos de los acontecimientos.

EXTRACTO 13.1

EL ENUNCIADO DEL PROBLEMA

A partir del estudio de las reacciones al *GI Bill*⁶ en los medios de comunicación, este artículo pretende explorar no sólo cómo, aparentemente, la educación universitaria se vuelve más accesible al americano medio –una posibilidad cargada con nuevo significado–, sino también lo que este fenómeno revela sobre los valores culturales subyacentes que informan sobre este cambio en la percepción de la universidad. De hecho, el *GI Bill* cambió la opinión de los estadounidenses sobre la educación en la universidad, y estas nuevas percepciones encajaban y estaban íntimamente unidas al nuevo concepto de uno mismo que se definía en un mundo corporativo y una cultura consumista. No obstante, estas nuevas concepciones también contienen nociones tradicionales del valor de la universidad, indicadores de clase social y, también, normas de género que ya existían antes pero en contraste con la asociación del acto con la democratización. Ningún historiador ha estudiado estas ramificaciones culturales más profundas relacionadas con el *GI Bill*. El enorme impacto del GI Bill ha sido considerado como un hecho. Consecuentemente, creo que no hemos estudiado adecuadamente el resultado más importante del *GI Bill* –su función en la remodelación del papel de la educación universitaria en la cultura de la posguerra de Estados Unidos.

Fuente: Clark, D. A. (1998), «The two Joes meet –Joe College, Joe Veteran: collage education, and postrar American culture» (p. 167-169). *History of Education Quarterly*, 38 (2), 165-189.

Una **fente secundaria** es el registro o testimonio de cualquiera que no haya sido testigo presencial o participante en el acontecimiento. Una fuente secundaria contiene la información proveniente de otra persona, que puede o no haber

⁶ Nota del revisor: El *GI Bill* es un proyecto de ley (Ley de veteranos) puesto en marcha tras la segunda guerra mundial en EE UU que introdujo un programa de financiación de estudios universitarios y de posgrado para los excombatientes y que sigue vigente en la actualidad como ayuda financiera a los soldados para realizar sus estudios.

presenciado el acontecimiento. Las fuentes secundarias contienen investigación histórica y sobre la elaboración de políticas que interpretan otras fuentes primarias y secundarias. Dichas fuentes proporcionan puntos de vista y posiblemente hechos para el análisis.

La clasificación de las fuentes en primarias o secundarias depende del problema de la investigación. Algunas fuentes pueden ser primarias en un estudio y secundarias en otro. El número de fuentes primarias necesario para un estudio varía según el tema. Para obtener fuentes primarias, el analista piensa en las fuentes que aportarían información al tema, y luego investiga qué documentos se conservan y si son accesibles. Un estudio puede manejar diferentes tipos de fuentes, pero es esencial que las fuentes primarias sirvan como base para la documentación. La documentación es el procesamiento de pruebas basado en cualquier tipo de fuente ya sea escrita, oral u objeto.

Localización de las fuentes. La credibilidad de un estudio analítico viene determinada, en gran parte, por las fuentes primarias seleccionadas. El enunciado del problema y las limitaciones indican las fuentes primarias necesarias. Un estudio sobre las normas de admisión de una universidad sería muy imperfecto sin documentos institucionales. Una biografía de Stanley Hall se pondría en duda si se ignoraran sus documentos y sus escritos privados. La bibliografía cita únicamente aquellas fuentes que se emplean en el estudio.

Los documentos. Los documentos se localizan a través de guías especializadas, catálogos, índices y bibliografías, o a través de centros de investigación. Algunos ejemplos de trabajos de referencia especializados son *A Catalogue of Rare and Valuable Early Schoolbooks* (Un catálogo de libros escolares raros y de gran valor), *Educational Periodicals During the Nineteenth Century* (Publicaciones educativas del siglo diecinueve) y *Selective and Critical Bibliography of Horace Mann* (Bibliografía selectiva y crítica de Horace Mann). Los investigadores pueden empezar con una búsqueda de documentos más amplia manejando guías de archivos y recopilaciones de manuscritos privados que proporcionan a menudo listas comentadas. El *A Guide to Manuscripts and Archives in the United States* (Guía de manuscritos y archivos en los EE UU) describe 1.300 depósitos y la *Guide to the National Archives of the United States* (Guía de archivos nacionales de EE UU) recoge un índice de registros educativos de agencias gubernamentales. Otros directorios son el *National Inventory of Documentary Sources in the United States* (Inventario Nacional de fuentes documentales en EE UU) y el *Directory of Archives and Manuscripts Repositories in the United States* (Directorio de depósitos de archivos y manuscritos de los EE UU). El *National Union Catalogue of Manuscripts Collection* (Catálogo de la unión nacional de recopilaciones de manuscritos), publicado anualmente por la Biblioteca del Congreso, cita el creciente número de recopilaciones educativas disponibles para los estudiosos. Los centros de investigación de archivos que

se encargan de temas históricos concretos cuentan frecuentemente con registros educativos⁷.

Los estudios sobre la elaboración de políticas educativas utilizan documentos gubernamentales y testimonios orales. Los documentos del gobierno federal están recogidos en un índice en el *Monthly Catalog of United States Government* (Catálogo mensual del gobierno de los Estados Unidos), el *Publication Referente File* (PRF) (Archivo de referencias de publicaciones) (PRF), el *Congressional Information Service/Index* (CIS) (Servicio/Índice de información del Congreso), el *American Statistical Index* (ASI) (Índice estadístico americano), el *Index to Government Documents* (Índice de documentos gubernamentales) y las listas de publicaciones de las agencias. El *Legislative Research Checklist* (Catálogo de investigación legislativa), que publica mensualmente el Consejo de Estado, a menudo resume los documentos publicados procedentes de comisiones o consejos que recomiendan y redactan la nueva legislación. Puesto que las herramientas de investigación para la historia legislativa estatal varían mucho de un estado a otro, se hace necesario consultar bibliotecas estatales sobre leyes o bibliotecas de referencias legislativas. Los documentos locales gubernamentales no están centralmente catalogados y pueden obtenerse a través de la agencia.

Los testimonios orales. Los testimonios orales que son relevantes para un tema requieren una previa planificación. El investigador decide qué personas conocen el tema, las localiza y recopila datos a partir de las entrevistas⁸. La selección de los informantes para los testimonios orales se realiza a través de técnicas de muestreo intencionado como el muestreo bola de nieve o la selección de casos atípicos (véase el capítulo 10). Un panel de expertos, por ejemplo, nombra a los individuos basándose en criterios a partir del enunciado del problema. Otros aspectos a tener en cuenta son la accesibilidad de los individuos y la viabilidad (tiempo, gastos, número de investigadores, etc.). El investigador establece los criterios de selección.

La historia oral está constituida por los datos recogidos mediante una entrevista guiada, similar a ciertas entrevistas en profundidad. El investigador desea información del asunto que investiga y cualquier otra información que pueda facilitarle la comprensión del tema. La obtención de testimonios orales lleva

⁷ Los archivos del *Labor History and Urban Affairs* en la Wayne State University son posiblemente la colección más antigua. Otros centros que tienen documentos relevantes para la historia de la educación son el *Urban Archives Center*, en la Universidad de Temple, los *Archives of the Industrial Society* de la Universidad de Pittsburg, los *Archives of the History of American Psychology*, en la Universidad de Akron, el *Ohio History of Education Project*, de la *Ohio Historical Society*, los *Social Welfare Archives*, en la Universidad de Minnesota y los *Television News Archives* de la Universidad de Vanderbilt.

Véase Hill. M. R. (1993), *Archivals strategies and techniques*.

⁸ Véase W. W. Cutler (1971).

mucho tiempo y acaban siendo extensas transcripciones. Las transcripciones están sujetas a crítica interna y externa como sucede con otras fuentes documentales usadas frecuentemente por este método. Por ejemplo, en un estudio de caso sobre el contexto de la elaboración de una legislación sobre educación bilingüe de Arizona (extracto 13.2), diecisiete personas fueron entrevistadas formalmente. Estas personas seleccionadas a través de estrategias de muestreo intencionado, se concentraban entre los «influyentes» en la política del estado, pero divididos en cuatro grupos (muy influyentes, influyentes, menos influyentes y participantes a menudo olvidados). Se utilizaron, además, como fuentes de datos los documentos de la Cámara y del Senado, del Ministerio de Educación, los reglamentos administrativos y dos periódicos.

EXTRACTO 13.2

SELECCIÓN DE TESTIMONIOS ORALES

El marco conceptual adoptado en este estudio fue desarrollado por Marshall, Mitchell y Wirt (1986) para identificar el «contexto de poder y la influencia de la elaboración de políticas» (p. 347-348). Este estudio sigue la pista del proceso que condujo a la promulgación de la ley sobre la educación bilingüe de 1984 en Arizona.

Los datos primarios que constituían este estudio se obtuvieron a partir de entrevistas con «la elite política» (Wirt, Mitchell y Marshall, 1985)... En total, 17 personas fueron entrevistadas formalmente durante la recogida de datos, algunos más de una vez. No obstante, los autores estuvieron implicados en grado diferente a lo largo de todo el proceso legislativo, asistieron a muchas reuniones y mantuvieron numerosos contactos informales y conversaciones relacionadas con la ley. Siguiendo la clasificación de grupos y ordenamiento de grupos políticos de influencia de Marshall y otros (1986, p. 355), las entrevistas se concentraron en los representantes de los grupos más cualificados de Arizona, pero divididos en cuatro grupos (centro, círculo cercano, círculo lejano y participantes a menudo olvidados).

Además, se analizaron las copias de todas las propuestas legislativas y enmiendas que llevaron a la versión de la ley finalmente adoptada. Arizona tiene una historia legislativa escrita breve, pero conserva un registro con todos los votos en el boletín de la Cámara y el Senado, que también incluye las declaraciones de algunos legisladores explicando sus votos finales. Los documentos del Ministerio de Educación, que incluyen reglamentos administrativos y propuestas legislativas relacionadas con la educación bilingüe, al igual que las ediciones anteriores del *Phoenix Republic* y el *Tucson Daily Star and Citizen*, fueron revisados para buscar artículos que hablaran de las deliberaciones y las acciones legislativas sobre la educación bilingüe.

Fuente: adaptado de Sacken, D. M. y Medina, M. Jr. (1990), «Investigating the context of state-level policy formation: A case study of Arizona's bilingual education legislation», *Educational Evaluation and Policy Analysis*, 12 (4), 391.

Crítica de las fuentes. Las técnicas de crítica interna y externa se aplican a todo tipo de fuentes, como documentos, testimonios orales y restos materiales. Incluso, las fuentes que son publicaciones oficiales o que se conservan en archivos están sujetas a la crítica. La crítica externa determina la autenticidad de la fuente. La crítica interna determina la credibilidad de los hechos declarados por la fuente. Aunque la crítica externa e interna hacen que nos planteemos diferentes preguntas sobre la fuente, las técnicas se aplican simultáneamente. La crítica de fuentes puede formar parte de una discusión metodológica o de un apéndice al estudio.

La crítica externa. La **crítica externa** determina si la fuente es un documento original, un documento falsificado o una variedad de un documento original. Las preguntas típicas son: «¿quién escribió el documento?» y «¿cuándo, dónde y cuál era la intención?». Cuanto más especializado sea el conocimiento del analista, más fácil le resultará determinar si un documento es o no original. El historiador necesita conocimientos sobre cómo vivía y se comportaba la gente de la época en que se elaboró el documento, sus creencias y cómo funcionaban sus instituciones. El investigador educativo tiene menos probabilidad de encontrarse con documentos falsos que el científico social que estudia movimientos políticos, religiosos o sociales controvertidos. No obstante, los títulos o las fechas de una institución pueden ser inexactos. A veces, resulta imposible determinar la contribución de un individuo a los informes gubernamentales o a discursos si existen diversos autores.

La fecha y el lugar en el que se escribió o publicó el documento pueden establecerse por medio de las citas del documento, la fecha de la recopilación de manuscritos o los contenidos del documento. No obstante, puede que los documentos de trabajo internos de una institución o los resúmenes redactados por una persona no contengan ninguna fecha o sean insuficientes para su utilización si sólo aparece el año.

Lo que el investigador educativo tiene más probabilidades de encontrar son fuentes discrepantes. Las fuentes discrepantes son dos o más textos del mismo documento, o dos o más testimonios del mismo acontecimiento. Por ejemplo, la relación de un periódico sobre los resultados de un programa estatal de evaluación educativa puede diferir del informe estadístico real publicado por el Ministerio de Educación y ambos pueden discrepar de documentos independientes del informe. En un caso así, el periódico, el informe oficial y los documentos independientes son fuentes auténticas de textos diferentes. Los testimonios orales de diferentes personas pueden ser fuentes auténticas pero discrepantes.

La crítica interna. La crítica interna determina la exactitud y el carácter fidedigno de las declaraciones en la fuente. El historiador se pregunta: «¿son exactas las declaraciones y son fiables los testimonios?». La exactitud está relacionada con la proximidad geográfica y cronológica del testigo al acontecimiento, con su competencia y con la atención prestada al acontecimiento.

Obviamente, no todos los testigos que presenciaron el suceso con la misma cercanía, son observadores que registran la información con el mismo nivel de competencia. Ésta depende de la experiencia, del estado de salud mental y física, del nivel educativo, de la memoria, de las habilidades narrativas, etc. De todos es conocido que un testigo presencial bajo condiciones traumáticas o estresantes tan sólo recuerda partes selectivas del acontecimiento, aunque esté convencido de que, por el mero hecho de estar presente, sus datos son exactos. Incluso si un testigo es competente, puede que su testimonio sea parte interesada o sesgada. El sesgo y los prejuicios hacen que el testigo normalmente distorsione, ignore o dé mucha más importancia de la necesaria a algunos incidentes. Las condiciones en las que se realizan las declaraciones pueden influir en su exactitud. El estilo literario, las leyes sobre difamación, las convenciones sobre el buen gusto o el deseo de agradar pueden llevar a un exceso de educación o de expresiones de aprecio.

Algunas técnicas calculan la exactitud y el carácter fidedigno de una declaración. Las declaraciones de un testigo que es indiferente ante lo que acaba de presenciar, las que perjudican a la persona que las enuncia o aquéllas que se oponen a sus deseos personales tienen menos probabilidad de estar sesgadas. Las declaraciones consideradas de conocimiento común o casuales tienen menos probabilidad de ser erróneas. Otras fuentes creíbles pueden confirmar, modificar o rechazar declaraciones. No obstante, en un análisis cualitativo, el hecho de que las declaraciones de testigos independientes coincidan puede resultar engañoso, ya que la investigación depende únicamente de las fuentes que se han conservado. El que coincidan con otros hechos conocidos o testimonios circunstanciales aumenta la credibilidad de una declaración. Un investigador hace referencia a la fuente cuando dice: «según la opinión del juez», «Horace Mann dice que» o «nos basamos en el Presidente de la Cámara cuando afirmamos que... ».

La crítica interna y externa requiere conocimiento sobre los individuos, los acontecimientos, el comportamiento de la gente del periodo estudiado. «La capacidad de ponerse en el lugar del otro... de interpretar documentos, acontecimientos y personalidades desde su punto de vista, criterios y gustos (sin dejar de lado los criterios propios) ha sido llamado a veces «mentalidad histórica» (Gottschalk, 1969, p. 136-137). A lo largo de todo el proceso, el analista es escéptico y crítico con las fuentes y las declaraciones. Un analista no queda satisfecho o convencido fácilmente de que las fuentes proporcionen testimonios lo más cercanos posible a los acontecimientos.

La crítica de las fuentes. La crítica de las fuentes se trata de distintas maneras dentro de un estudio. Las citas más obvias son las notas a pie de página y las partes denominadas *Notas*, *Referencias* o *Bibliografía* que siguen al estudio. La lista, normalmente, se puede prolongar. Por ejemplo, el artículo de revista de 12 páginas «*A History of Discrimination Against Black Students in Chicago Secondary Schools*» (Una historia de la discriminación contra los estudiantes negros en los

institutos de Chicago) (Daniel, 1980) posee 83 notas a pie de página, que incluyen informes oficiales del Ministerio de Salud, de Educación y de Bienestar de Estados Unidos, junto con los de otros departamentos federales y agencias, legislación de Illinois, informes anuales sobre el Consejo de Educación, el Censo de los Estados Unidos, artículos del *Chicago Tribune* y del *Chicago Defender*, el Código municipal de Chicago, entrevistas con los participantes, informes de la Comisión de Chicago sobre las relaciones raciales, la vicecomisión de Chicago y el *Chicago Real State Board*, así como fuentes secundarias.

En las discusiones metodológicas del estudio o en un apéndice, el historiador también hace referencia a la crítica. El autor del extracto 13.3 critica las fuentes al hacer notar que las cartas largas, las breves respuestas de los cuestionarios de la década de 1890 y la información aportada por las alumnas de Troy deben «interpretarse cuidadosamente». El investigador señala el posible sesgo de los documentos debido a la autoselección de las que acudieron a Troy, y porque las que mantuvieron contacto con Troy probablemente tuvieron experiencias favorables en el colegio y desarrollaron valores feministas.

EXTRACTO 13.3 **CRÍTICA DE LAS FUENTES**

Un grupo de alumnas de Troy manifestó su opinión acerca de la influencia que, a largo plazo, había tenido sobre ellas la institución y su fundadora. Algunas de las que respondieron a los cuestionarios de la década de 1890, unas con largas cartas mientras que otras con breves respuestas, aportaron alguna luz sobre cómo recordaban la experiencia. Y finalmente, otras, simplemente describiendo sus vidas, inadvertidamente aportaron testimonios sobre el tipo de carácter que Troy reforzaba. Lo que más destaca en la mayor parte de los documentos es la enorme importancia de la personalidad de Willard, que proporcionaba a sus alumnas una nueva imagen de lo que las mujeres podían llegar a ser...

Evidentemente, dichos datos deben interpretarse con cautela. Obviamente, es muy probable que una mujer que se sienta muy unida a Troy escriba una respuesta detallada en el cuestionario; y también es probable que las niñas que recuerdan tal unión se hayan tomado la molestia de contestar. Hubo, ciertamente un proceso de autoselección en el grupo de las mujeres que escogieron ir a Troy. Sin embargo, estos fragmentos conducen a testimonios del principio de un nuevo tipo de personalidad, la mujer educada que no se avergonzó de aprender y que, inevitablemente, tendría una noción más amplia de lo que ofrece el mundo, que sus hermanas a las que no se había animado a leer o a pensar por sí mismas. (35)

Fuente: «The Ever Widening Circle: The Difusión of Feminist Values from the Troy Female Seminary, 1822-1872», de Anne Firor Scott, *History of Education Quarterly*, vol. 19, primavera de 1979. Reimpreso con el permiso del autor.

Los apéndices metodológicos se escriben porque el estudio utiliza fuentes que se recopilan y analizan de nuevo. Un apéndice, más que el estudio *per se*, permite que el analista se explaye en los problemas metodológicos de las fuentes y que comente su valor académico. El extracto 13.4 ilustra una referencia a problemas metodológicos y la necesidad de un apéndice metodológico. Este estudio requería la identificación de todas las huelgas del profesorado que tuvieron lugar entre 1978 y 1979. La información recogida por las agencias gubernamentales estatales y nacionales y por las asociaciones profesionales estatales y nacionales contenía «graves discrepancias» a causa de las distintas definiciones de huelga y de lagunas de información de las fuentes. El método de identificación de las 158 huelgas se presentó en el apéndice. Así, el método resulta accesible a otros investigadores.

EXTRACTO 13.4 **CRÍTICA DE LAS FUENTES**

Una tarea de investigación muy importante consistía en localizar y observar las huelgas del profesorado que tuvieron lugar durante 1978 y 1979. Necesitábamos identificar todos los lugares de las huelgas para poder llevar a cabo una encuesta por correo de los distritos afectados y necesitábamos localizar los lugares en los que se podrían realizar los estudios de campo. Conforme avanzábamos, la tarea de identificar las huelgas se volvía extraordinariamente difícil. No hay ninguna agencia nacional central que disponga de un sistema fiable para identificar rápidamente los lugares de las huelgas. La información recopilada por las asociaciones profesionales estatales y nacionales y por las agencias gubernamentales estatales y nacionales, contiene graves discrepancias. Algunas se deben a que las definiciones de lo que es una huelga son diferentes. Otras, a las lagunas de información de las fuentes. La base de nuestro cálculo de que hubo 158 huelgas se describe en un apéndice técnico...

Fuente: Colton, D. L. y Graber, E. E. (1980), *Enjoining teacher strikes: The irreparable harm standard* Concesión N° NIE-G-78-0149, 26. Washington, DC: National Institute of Education.

Hechos, generalizaciones e interpretaciones

Los **hechos** son la base de las generalizaciones que pueden ser interpretadas como explicaciones causales. Sin embargo, el proceso no es sencillo. La crítica de las fuentes puede conducir a la reformulación del problema y a una búsqueda más profunda de fuentes y hechos.

1. *Hechos que describen el quién, qué, cuándo y dónde de un acontecimiento.* Sin embargo, la mayoría de los analistas van más allá, obteniendo hechos

descriptivos y planteándose preguntas que interpretan la manera en que un acontecimiento histórico tuvo lugar y las razones para que sucediera. Así pues, el analista va de la identificación de los hechos al enunciado de generalizaciones para extraer explicaciones causales. Las preguntas realizadas a las fuentes son cruciales a lo largo del proceso.

La destreza del investigador en la realización de preguntas es similar a la del detective en la búsqueda de pruebas y a la del científico que sistemáticamente pone a prueba dicha evidencia. Las preguntas pueden ser muy específicas, como: «¿cuándo murió Henry Barnard?» o pueden ser abstractas, como por ejemplo, «¿cómo influyó el progreso científico en las prácticas de la administración escolar?». La formación metodológica y la experiencia, el conocimiento general y especializado, la intuición disciplinada y la lógica influyen en el análisis. El investigador emplea, básicamente, un método de razonamiento inductivo, yendo desde los hechos específicos hasta las generalizaciones. Cuantas más preguntas se formulen a las fuentes sobre el tema, más complejo y comprensible será el análisis.

Cuando las declaraciones y los hechos se contradicen, la información adicional intenta resolver las aparentes diferencias. Sin embargo, eventualmente, el investigador deberá tomar una decisión. Incluso la frase «El colegio femenino de Troy, abierto oficialmente en 1821, pero cuyas raíces comenzaron en 1814, fue la primera institución permanente en ofrecer a las mujeres estadounidenses un currículum similar al de sus compañeros varones contemporáneos.» (Scott, 1979, p. 3) requirió una «prueba decisiva», prueba que confirma un punto de vista y rechaza los demás. La decisión de aceptar una declaración como un hecho reside en su *credibilidad* y su *posibilidad* (Barzun y Graff, 1992). En otras palabras, dada una serie de hechos, es razonable asumir que el acontecimiento ocurrió de una cierta manera, o que no sucedió así. Los hechos se pesan y juzgan por su consistencia, la acumulación de pruebas y otras técnicas.

2. Las interpretaciones de las relaciones entre los hechos son generalizaciones. Cuando D. M. Potter escribe «la generalización en la historia es inevitable y... el historiador no puede evitarla haciendo declaraciones limitadas sobre datos limitados. Para un microcosmos es tan infinito como para un macrocosmos. Y no sólo eso, sino que, además, las relaciones entre los factores en un microcosmos son tan sutiles y las generalizaciones vinculadas con la declaración de tales relaciones son tan amplias como las generalizaciones de las relaciones entre los factores de una situación a gran escala» (Potter, 1963, p. 191). Cada generalización está sujeta a un análisis y, normalmente, se modifica o acepta. Los elementos que, a menudo, aparecen como hechos en un estudio son frecuentemente generalizaciones de los hechos para los que no se dispone de espacio para su presentación en el estudio. El extracto 13.5, por ejemplo, tomado de un texto de un estudio histórico, contiene hechos (la señorita Strachan fue presidenta de la Asociación intermunicipal de profesoras, 1906-1912), generalizaciones

EXTRACTO 13.5

EJEMPLO DE GENERALIZACIONES

Quizá, la razón más importante por la cual empezó el movimiento de igualdad de salarios en Nueva York pueda encontrarse en una de aquellas afortunadas coincidencias históricas en las que una sola persona, equipada con el carácter adecuado, cualidades de líder, ambición e interés, se encuentra en el lugar apropiado en el momento oportuno. Sin duda, nadie puede estudiar la historia del movimiento de igualdad de salarios sin concluir que habría sido muy diferente y, probablemente, mucho menos efectivo si Grace Strachan no hubiera surgido como su líder y portavoz de las mujeres, igual que el más irritante de los tábanos del *establishment* educativo y político. Procedente de la Escuela Normal de Búfalo, la señorita Strachan llegó a las escuelas de Nueva York como maestra, probablemente a finales de la década de 1880. En el momento en el que el movimiento de igualdad de sueldos estaba en su apogeo, ella trabajaba como superintendente de distrito en los distritos 33 y 35, que contaban con más de 32.000 niños escolarizados. Una de las organizadoras de la Asociación intermunicipal de profesoras en 1906, la señorita Strachan, se convirtió en su presidenta en 1907 y ocupó tal puesto hasta 1912, cuando, a efectos prácticos, el conflicto había terminado.

Fuente: «Tempest on the Hudson: The struggle for "Equal Pay for Equal Work" in the New York City Schools, 1907-1911» de Robert E. Doherty, *History of Education Quarterly*, vol. 19, invierno de 1979. Reimpreso con el permiso del autor.

(«cuando el conflicto había terminado», «la portavoz de las mujeres», «los más irritantes de los tábanos») y causas de un acontecimiento (un líder, una organización, salarios desiguales entre hombres y mujeres) y esto implica dos teorías de cambio: circunstancia (un individuo con cualidades determinadas «se encuentra en el lugar adecuado en el momento oportuno») y el «líder» («carácter adecuado, cualidades de líder, ambición e interés»).

3. *Las explicaciones analíticas son síntesis abstractas de generalizaciones, presentadas normalmente como conclusiones.* Las generalizaciones presentadas a lo largo de todo el estudio se vuelven a analizar para su contextualización, consistencia interna, documentación, acumulación de pruebas e inducción lógica. El proceso es cíclico, se vuelve constantemente a los hechos y, si resulta necesario, a los documentos para deducir el sentido. Una explicación causal que aparece como una visión de conjunto en una introducción no significa que el investigador parta de esta interpretación y que pretenda demostrar sus conocimientos personales. La visión global introductoria ha sido probablemente la última sección del estudio en escribirse, porque la lógica del estudio debe surgir de ella y los criterios para juzgar su calidad derivan de aquélla.

Las conclusiones son generalizaciones sintetizadas y conceptos previamente documentados en el estudio. En otras palabras, las conclusiones son un resumen

interpretativo de las generalizaciones. Las conclusiones pueden redactarse en forma narrativa o como un breve listado, seguidas por afirmaciones sobre el estado del conocimiento sobre el tema, la identificación de las consecuencias políticas o las sugerencias para otras investigaciones. El extracto 13.6 ilustra una conclusión que contiene tres interpretaciones (párrafo 1) de las maneras en que se utiliza la ciencia social en la elaboración de política legislativa, las razones por las que se utiliza y los efectos de su utilización. Se hacen dos recomendaciones políticas (párrafo 2) a partir de estos descubrimientos.

LA INVESTIGACIÓN ANALÍTICA EN PERSPECTIVA

Los estudios analíticos sobre temas educativos ayudan al desarrollo del conocimiento y a la mejora de las prácticas. A través de la comprensión de los conceptos educativos y los acontecimientos del pasado, se pueden entender mejor las políticas, las tendencias y prácticas educativas del presente. A continuación presentamos algunos empleos de la investigación analítica.

1. *El análisis de conceptos educativos puede ayudar en la selección de problemas de investigación, planes y metodología.* El análisis del concepto de enseñanza como diferentes tipos de acciones educativas, por ejemplo, puede sugerir preguntas de investigación y ayudar en los planes de otras investigaciones, ya sean cuantitativas o cualitativas, sobre la enseñanza. Un mapa conceptual del dominio lógico de la enseñanza puede exponer una mejor comprensión del tema central de la investigación previa y sugerir temas que necesitan más investigación complementaria. Un análisis de un concepto puede proporcionar intuiciones para la selección de un proyecto. Conceptos tales como *objetivos conductuales, escolarización alternativa y solución de problemas* no son simples palabras, sino que forman parte de un sistema lingüístico que construye un marco para organizar ideas en la mente del investigador. Aclarar el significado del concepto antes de la recopilación de datos, permite al investigador seleccionar variables y medidas más apropiadas para investigar el concepto. En la investigación cualitativa, el análisis de conceptos es, a menudo, el centro exclusivo del estudio o bien el análisis puede identificar categorías para organizar e interpretar los datos.

2. *La investigación analítica proporciona conocimientos sobre los fundamentos de ideas, instituciones, líderes, políticas y prácticas educativas.* El conocimiento del pasado informa a los profesionales de la educación, a los elaboradores de políticas y a la sociedad en general sobre la educación y su papel en la sociedad estadounidense. Este conocimiento, a menudo, cuestiona los mitos de la educación. Al examinar la mayor o menor fortuna de soluciones adoptadas en el pasado frente a problemas que aún persisten, los que deben

EXTRACTO 13.6**LAS CONCLUSIONES ANALÍTICAS Y SUS IMPLICACIONES****CONCLUSIÓN**

Los datos aquí analizados apoyan tres conclusiones básicas relativas a la utilización de la ciencia social dentro de las legislaturas estatales: (1) la ciencia social es sólo uno de los cuatro tipos diferentes de autoridad experta influyente en las decisiones legislativas. Los otros tipos de experiencia (legal, política y técnica) son más frecuentes y, por lo tanto, se utilizan más a menudo. No obstante, los legisladores a tiempo completo y los asesores profesionales recurren frecuentemente a las ciencias sociales. (2) La utilización de las ciencias sociales presta una contribución en los aspectos intelectuales y sociales del proceso de toma de decisiones. Intelectualmente, la experiencia científica orienta a los elaboradores de políticas hacia ciertos aspectos de las decisiones que deben ser adoptadas. Socialmente, la utilización de la ciencia altera los procesos de colaboración y conflicto que controlan las consecuencias de las decisiones. (3) Las contribuciones de la ciencia social cambian significativamente conforme las propuestas legislativas evolucionan a lo largo de las cuatro fases de la toma de decisiones: articulación, reunión, asignación y examen. Durante la primera fase, la ciencia puede influir de manera poderosa y exenta de polémica. Conforme los resultados legislativos se definen de forma más clara, y se especifica quiénes saldrán beneficiados y perjudicados por la política propuesta, la utilización de la ciencia se vuelve más partidista y más polémica, y se transforma, por tanto, en una herramienta o un arma de conflicto político.

Estos resultados sugieren que los científicos sociales, los legisladores estatales y las agencias patrocinadoras serán útiles si: (a) el desarrollo del apoyo científico a varias soluciones políticas se reconoce como un proceso *político*, que requiere de la contribución de personas que se declaren abiertamente partidistas al igual que de profesionales científicamente sofisticados y si (b) se realizan esfuerzos sistemáticos para que los análisis científicos aparezcan *antes* en el proceso legislativo. Es decir, la utilización de la ciencia debe verse como una movilización de la experiencia científica a favor de intereses políticos específicos de legisladores estatales que deben tomar decisiones políticas más que científicas. La utilización de la ciencia sirve mejor para perfeccionar y criticar --no para reemplazar-- otros mecanismos más importantes para la definición y resolución de asuntos políticos públicos.

NOTA: Beca el *National Institute of Education* N° NIE-G-76-0104 titulado «Improving Social Science Utilization in Legislative Policy Making for Basic Skills Education». Esta beca a la Universidad de California, Riverside (Douglas E. Mitchell, investigador principal) fue concedida en octubre de 1976.

Fuente: «Social Science Impact on Legislative Decision Making: Process & Substance» de Douglas E. Mitchell, *Educational Researcher*, noviembre de 1980. Copyright 1980. American Educational Research Association. Reimpreso con permiso.

tomar decisiones, se vuelven más realistas y moderados en sus peticiones y están más informados en sus elecciones.

3. *La investigación analítica puede aclarar las actuales discusiones jurídicas y políticas al interpretar el pasado con una distancia objetiva y un juicio histórico razonado.* Los estudios analíticos interpretan la complejidad y las relaciones colectivas educativas, sociales, económicas, legales y políticas del pasado. F. M. Wirt sugiere que dicha investigación resulta útil porque «precisamente gracias al proceso de hacer desaparecer todo lo que no está corroborado por datos, la política pública se enriquece» (1980, p. 17). La investigación analítica, sin embargo, nunca afirma predecir futuras acciones educativas y políticas.

4. *La investigación histórica puede desmitificar las nociones idealizadas sobre los acontecimientos pasados y estimular la acción.* La investigación histórica ha documentado prácticas y políticas discriminatorias en la educación, algunas de las cuales han estado institucionalizadas desde el periodo colonial, y apoyadas por los valores sociales y culturales de la época. Algunos investigadores revisionistas se han cuestionado los estereotipos asignados a algunos inmigrantes y grupos étnicos. Otros investigadores reconstruccionistas se preguntan por la ausencia, en los libros de texto y otros materiales educativos, de la contribución a la nación de las mujeres y de las minorías.

5. *La investigación histórica, en un sentido más amplio o quizá más filosófico, puede crear un sentido de objetivo común.* Las interpretaciones históricas reflejan la opinión fundamental de que la educación pública en Estados Unidos ha servido y puede servir al bien común. Los logros educativos y el papel de la educación en la sociedad estadounidense se menosprecian, a menudo, en la investigación empírica fragmentada. En el objetivo de la investigación analítica está implícito el interés de que los éxitos en educación y las prácticas educativas benefician tanto a los individuos como a la sociedad estadounidense.

6. *La investigación histórica es un área dinámica de indagación educativa porque cada generación reinterpreta su pasado.* Los filósofos analíticos se centran en los significados cambiantes de los conceptos educativos. Los historiadores educativos, especialmente los revisionistas, formulan nuevas preguntas, utilizan una mayor variedad de fuentes, analizan el pasado a través de un amplio conjunto de conceptos de la ciencia social y aplican procedimientos cuantitativos cuando resultan necesarios. A menudo, se revisan las conclusiones de los estudios históricos y políticos. Un acontecimiento histórico reciente, una sentencia de un tribunal o una nueva política pueden indicar un cambio básico en los principios, políticas y éxitos educativos colectivos previos.

LOS PATRONES DE CREDIBILIDAD PARA LOS ESTUDIOS HISTÓRICOS

La investigación analítica requiere procedimientos metodológicos para expresar un tema analítico, localizar y criticar fuentes primarias, establecer hechos y

realizar generalizaciones sobre explicaciones o principios causales. Estos procesos de investigación proporcionan criterios para juzgar un estudio histórico o la elaboración de políticas, como investigación verosímil. Los criterios que juzgan la adecuación de estudios históricos están seguidos por criterios para evaluar la investigación jurídica.

El lector juzga un estudio según la relación lógica entre el enunciado del problema, las fuentes, las generalizaciones y las explicaciones causales. La lógica del estudio completo proviene del enunciado del problema. En la evaluación de un estudio está implícita la pregunta: «¿ha alcanzado el analista el objetivo que se había propuesto?». Si todos los elementos de la investigación no aparecen explícitos, el estudio puede ser criticado por sus desviaciones o por mantener conclusiones sin justificación.

Los *enunciados del problema* en la introducción delimitan el estudio y se evalúan a partir de las siguientes preguntas.

1. ¿El tema es adecuado para la investigación analítica, es decir, se centra en el pasado o en el pasado reciente?
2. ¿El enunciado del problema indica claramente la información que se incluirá y la que se excluirá del estudio?
3. ¿Se expone el marco o la visión analítica?

La *selección y crítica de las fuentes* se evalúan según la relevancia para el enunciado del problema. Las fuentes aparecen enumeradas en la bibliografía y notas a pie de página. La crítica de las fuentes puede plantearse en el estudio, en las notas a pie de página o en un apéndice metodológico.

1. ¿Utiliza el estudio fuentes primarias significativas para el tema?
2. ¿Se exponen los criterios para la selección de las fuentes primarias?
3. ¿Se utilizaron fuentes auténticas para la documentación?
4. ¿Expone el analista una crítica de las fuentes?

Los *hechos y las generalizaciones* presentados en el texto se evalúan a través de las siguientes preguntas:

1. ¿Indica el estudio la aplicación de críticas externas para descubrir los hechos? Si aparecen hechos contradictorios, ¿se ofrece alguna explicación razonable?
2. ¿Las generalizaciones son razonables y están relacionadas lógicamente con los hechos?
3. ¿Las generalizaciones son apropiadas para el tipo de análisis? Por ejemplo, uno esperaría que hubiera una generalización mínima en un estudio que reproduce una serie de documentos en su texto original o incluye una serie de declaraciones políticas en orden cronológico. Se esperaría alguna síntesis en un análisis descriptivo o comparativo.
4. ¿Las generalizaciones se califican o formulan de manera provisional?

Las interpretaciones presentadas como conclusiones, son evaluadas según los siguientes criterios:

1. ¿Son razonables las interpretaciones y están relacionadas de manera lógica con los hechos y las generalizaciones presentados en el estudio?
2. ¿Sugieren las explicaciones múltiples causas para acontecimientos humanos complejos?
3. ¿Responde el estudio a todas las preguntas presentadas en la introducción, es decir, cumple el propósito del estudio?

RESUMEN

Los puntos siguientes resumen las características más importantes de la metodología analítica y su aplicación en los estudios educativos conceptuales, históricos, de elaboración de políticas y legales.

1. La investigación analítica, a diferencia de la investigación experimental, describe e interpreta el pasado o pasado reciente a partir de fuentes seleccionadas.
2. Las fuentes son documentos escritos, testimonios orales y restos materiales.
3. Las fuentes primarias son documentos o testimonios de testigos de un acontecimiento. Las fuentes secundarias son documentos o testimonios de personas que no presenciaron directamente el acontecimiento.
4. Los estudios analíticos explican generalizaciones de hechos (quién, qué, dónde y cuándo), sobre un acontecimiento y establecen interpretaciones que proponen múltiples causas para un único hecho.
5. Un análisis conceptual se centra en los significados del lenguaje de la educación por medio de la descripción del significado genérico, los distintos significados y el uso apropiado del concepto.
6. Los temas históricos se centran en biografías, movimientos, instituciones y prácticas. Los temas políticos incluyen un análisis de la distribución de poder entre los implicados en el sistema, los procesos de elaboración de políticas y el contenido político, desde la formulación hasta la promulgación.
7. Las biografías interpretativas reconocen el papel del investigador dentro del estudio.
8. Un problema histórico se encuentra delimitado por un periodo de tiempo, la localización geográfica, el acontecimiento estudiado específico y el punto de vista del análisis.
9. Las bibliografías e índices especializados localizan las fuentes primarias necesarias para la investigación histórica.

10. Los testimonios orales son entrevistas cualitativas en profundidad e informativas para estudiar acontecimientos pasados o recientes.
11. La crítica externa determina si la fuente es un documento original, un documento falsificado o una variante del documento original. La crítica interna determina la exactitud y la credibilidad de la información obtenida de la fuente.
12. La investigación histórica proporciona conocimientos y explicaciones sobre el pasado, clarifica las discusiones políticas presentes al interpretar el pasado con distanciamiento objetivo, revisa los mitos del pasado y puede crear un sentimiento de objetivo común hacia la educación en la sociedad estadounidense.
13. Los estudios históricos no pretenden predecir acontecimientos futuros de una manera exacta.
14. Los patrones de credibilidad para la investigación histórica enfatizan la relación lógica entre el enunciado del problema, la selección y crítica de las fuentes, por un lado, y los hechos, las generalizaciones y las explicaciones, por otro.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La investigación analítica difiere de la experimental en que...
 - a. utiliza principalmente documentos que han sido conservados.
 - b. se centra en los acontecimientos pasados o recientes.
 - c. emplea la inducción lógica para interpretar hechos.
 - d. todo lo anterior es correcto.
2. Las explicaciones interpretativas...
 - a. normalmente son generalizaciones sesgadas de los hechos.
 - b. sugieren causas simples para acontecimientos educativos.
 - c. son síntesis de generalizaciones que resumen hechos relacionados.
 - d. no son revisadas muy a menudo en estudios posteriores.
3. Un análisis conceptual de un concepto educativo se realiza para identificar...
 - a. sólo el significado esencial de un concepto.
 - b. los ejemplos y contraejemplos de los usos típicos del concepto a través de muestreo aleatorio.
 - c. los significados no típicos de un concepto.
 - d. el significado genérico, los significados diferentes y el uso apropiado del concepto.
4. ¿De las siguientes preguntas, cuál es una pregunta histórica?

- a. ¿Hay alguna relación entre la variable x y la variable y ?
 - b. ¿Hay una diferencia estadísticamente significativa entre los resultados de un grupo experimental y de control?
 - c. ¿Cómo se burocratizaron las escuelas durante la década de 1920?
 - d. ¿Cuál es el nivel de desarrollo cognitivo de los estudiantes en un programa de aceleración de matemáticas de quinto curso?
5. Ordene los siguientes pasos para llevar a cabo un estudio histórico:
- a. Identificar los hechos y obtener colaboración.
 - b. Solicitar técnicas de crítica de las fuentes.
 - c. Usar bibliografías e índices especializados para localizar las fuentes.
 - d. Adquirir el conocimiento de los antecedentes para el enunciado del problema.
 - e. Seleccionar un tema de investigación.
6. La crítica de las fuentes requiere:
- a. Cuestionarse la exactitud de cada afirmación de una fuente.
 - b. Cuestionarse la autenticidad de cada fuente.
 - c. El uso de conocimientos especializados sobre la manera en que la gente vivía y se comportaba en aquella época.
 - d. Todas las opciones anteriores son correctas.
- 7-9. Empareje los siguientes estudios con los tipos de investigación analítica correspondientes:
7. Un estudio sobre el proceso de toma de decisiones de una junta de dirección de una escuela. a. investigación histórica.
 8. Una biografía del fundador de la Asociación Nacional de Educación. b. investigación política.
 9. Un análisis de los significados de educación progresiva. c. análisis de un concepto educativo.
10. Los estudios analíticos cumplen varias funciones en la investigación educativa. ¿Cuáles son las funciones que los estudios analíticos *no* cumplen tan adecuadamente?
- a. Aclarar los significados de la educación que se utilizaron, de forma colectiva, en el pasado y, tal vez, en los debates políticos presentes.
 - b. Proporcionar conocimiento e interpretaciones de hechos históricos educativos y políticos del pasado.
 - c. Predecir de manera exacta los acontecimientos futuros.
 - d. Identificar el sentido de herencia y objetivo comunes en la educación estadounidense.

Problemas aplicados

Conteste las siguientes preguntas para cada problema metodológico.

1. Un investigador político quiere estudiar la toma de decisiones de una escuela local.
 - a. ¿Cómo podría delimitarse y centrarse este tema?
 - b. Enumere los tipos de fuentes posibles y ejemplos de fuentes para este tema.
2. Un investigador quiere estudiar la disciplina de los alumnos.
 - a. ¿Cómo podría enunciarse el problema de investigación si fuera un estudio histórico?, ¿y si se tratase de un estudio jurídico?, ¿y, un estudio político?, ¿y, un análisis de un concepto?
 - b. Establezca, al menos, una bibliografía o índice especializado para cada tipo de estudio.
3. Un investigador está estudiando la vida del Dr. Henry Daniel, que trabajó como director de una escuela estatal desde 1959 hasta 1979. El siguiente artículo apareció en un periódico e informaba sobre los discursos de varios asistentes a la cena de homenaje al Dr. Daniel tras veinte años de servicio como superintendente estatal de educación.

Más de cien altos cargos del sistema educativo de todo el estado honraron anoche al Dr. Henry Daniel en el Hotel Johnson de la capital del estado. Siguiendo las sugerencias de varios directores, se le entregó una placa grabada en reconocimiento de su destacada jefatura en la educación pública.

El gobernador señaló que gracias a los esfuerzos del Dr. Daniel, el estado ha adoptado un sistema de bachillerato que ha permitido cubrir las necesidades estatales relacionadas con la educación técnico-vocacional para la industria del estado, proporcionando las únicas instituciones de educación superior en regiones rurales y proporcionando una mejor educación general a los estudiantes de primer y segundo año que los institutos y universidades con estudios de cuatro años de duración.

El presidente de la organización de profesores elogió al Dr. Daniel por sus esfuerzos para aumentar los sueldos de los profesores de las escuelas públicas y para mantener la profesionalidad al elevar los requisitos para la titulación de los maestros. No obstante, el presidente señaló que los sueldos de los profesores de las escuelas públicas continúan siendo más bajos que la media nacional.

El presidente de la Asociación estatal para el desarrollo y la supervisión curricular declaró que los esfuerzos del Dr. Daniel han hecho posible el programa de evaluación de competencias mínimas. Esta innovación ha elevado el nivel de las asignaturas de bachillerato y ha demostrado al público que el diploma de bachillerato representa un alto nivel de «competencia educativa».

- a. ¿Por qué razón dudaría el investigador sobre la exactitud de lo afirmado en este documento?
- b. ¿Cómo podría el investigador corroborar la información del documento?

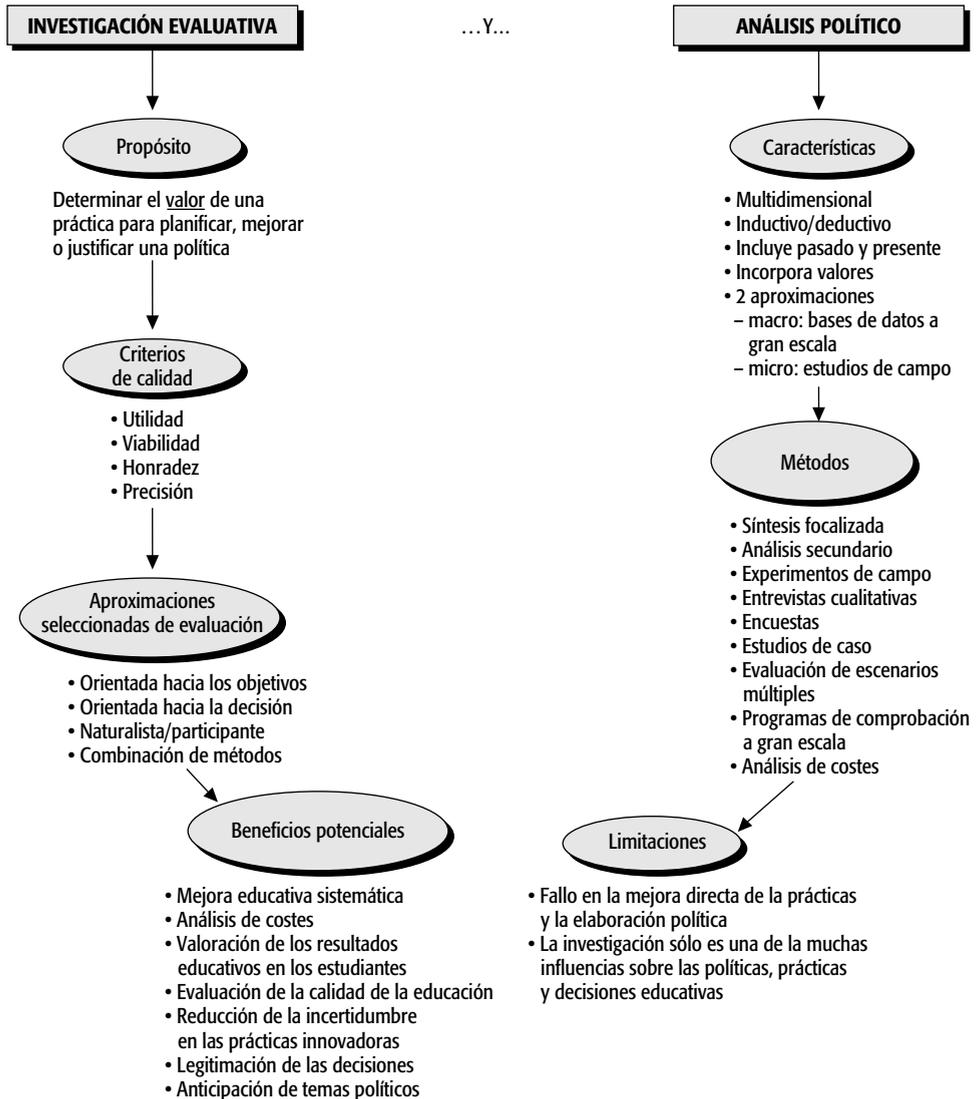
V

Diseños y métodos de investigación política y evaluativa

La mayor parte de los educadores reconocen la creciente importancia de la investigación sobre evaluación formal así como sobre análisis de planes. Frecuentemente, los educadores se preguntan cómo la evaluación y la investigación política pueden ayudar a la gente que toma decisiones y a distintos grupos elaboradores de políticas a mejorar nuestras escuelas y concebir planes educativas correctas. ¿Cuáles son las similitudes y las diferencias que existen entre evaluar un programa y evaluar una política? ¿Cómo se decide qué diseños y métodos son más apropiados a la hora de emplearlos en una situación determinada? ¿Pueden la investigación evaluativa y el análisis político proporcionar también más conocimiento educativo general sobre prácticas específicas comunes a muchas escuelas?

En la parte V describiremos los diversos enfoques que se emplean en la evaluación de las prácticas educativas. Los diferentes planteamientos destacan distintas preguntas relacionadas con una práctica específica. Algunas evaluaciones y análisis políticos son principalmente cuantitativos, otros son principalmente cualitativos y otros pueden combinar ambos métodos, el cuantitativo y el cualitativo.

Investigación evaluativa y análisis político



PALABRAS CLAVE

evaluación	evaluación respondente
valor	audiencias
evaluación formativa	combinación de métodos
evaluación sumativa	análisis político
enfoque evaluativo	evaluación de escenarios múltiples
evaluación orientada hacia los objetivos	análisis coste-beneficio
grupo objeto	análisis coste-efectividad
objetivos conductuales	análisis coste-utilidad
evaluación orientada a la decisión	análisis coste-viabilidad
evaluación naturalista/participante	

Este capítulo gira sobre los capítulos anteriores que hablaban sobre el diseño, la recopilación de datos y el análisis de datos. Los evaluadores y analistas políticos necesitan amplios conocimientos de investigación para tomar decisiones con suficiente información a lo largo de todo el método experimental; los educadores necesitan conocimientos de investigación suficientes para encargar un estudio e interpretar un informe de evaluación o un informe político.

La evaluación es la aplicación de las habilidades de investigación para determinar el valor de la práctica educativa. La investigación evaluativa ayuda a tomar decisiones en un lugar(es) determinado(s) e incrementa el conocimiento científico disponible sobre una práctica específica que, a menudo, es relevante para audiencias más generales. Las decisiones son aquellas que planifican y mejoran una práctica o que justifican (o no) la adopción extendida de una práctica. Un evaluador es a la vez un investigador y un educador cuyo trabajo resulta esencial para el funcionamiento global de las organizaciones educativas.

El análisis político evalúa los programas de gobierno para proporcionar a sus creadores recomendaciones pragmáticas entre alternativas políticas. Normalmente, se tienen en cuenta los efectos potenciales económicos y/ o políticos. Los dos enfoques en un análisis político son un macro-enfoque, que utiliza bases de datos amplias, y un micro-enfoque, que es una investigación ecléctica orientada al campo concreto. Un programa puede considerarse como un recurso específico adoptado para llevar a cabo una política.

En este capítulo resumiremos los propósitos de la investigación evaluativa, presentaremos una visión general de los enfoques de evaluación y hablaremos de los tres enfoques: orientado hacia los objetivos, orientado hacia las decisiones, y orientado de manera naturalista o participante. Las características y métodos del análisis político, incluyendo el análisis de costes, también quedan resumidas. También citaremos los patrones, los usos potenciales y las limitaciones de la investigación evaluativa y del análisis político.

PROPÓSITOS Y DEFINICIÓN DE LA INVESTIGACIÓN EVALUATIVA

Las actividades de evaluación siempre han formado parte de la educación. Frecuentemente, los juicios profesionales se han aplicado a la asignación de estudiantes a programas especiales, al aprendizaje del estudiante, la selección de materiales y la modificación de programas. En el pasado, dichas actividades se realizaban, a menudo, de manera informal y no sistemática. La necesidad de evaluaciones formales aumentó cuando la sociedad empezó a asignar mayores responsabilidades y recursos a la educación. La investigación evaluativa se usaba frecuentemente con el objetivo de rendir cuentas. Hoy en día, los estudios de evaluación se utilizan para determinar la asignación de recursos escasos y la eficacia de programas educativos alternativos, y para hacer posible y justificar decisiones valorativas en muchos aspectos de la educación.

Propósitos y roles de la evaluación

Una breve definición de la investigación evaluativa es la determinación del *valor* de un programa, un producto, un procedimiento o un objetivo educativo, o de la posible utilidad de enfoques alternativos para conseguir propósitos específicos. Las tres razones principales para llevar a cabo la investigación evaluativa son la *planificación*, la *mejora* y la *justificación* (o no justificación) de procedimientos, programas y/o productos.

La mayor parte de los educadores reconocen que la evaluación puede utilizarse con un objetivo formativo (como mejorar un programa) o un objetivo sumativo (como decidir si ese programa debe mantenerse). Más específicamente, la evaluación puede:

1. Ayudar a la planificación para el establecimiento de un programa.
2. Ayudar a tomar una decisión sobre la modificación de un programa.
3. Ayudar a tomar una decisión sobre la continuación o desarrollo de un programa.
4. Obtener pruebas para apoyar u oponerse a un programa.
5. Contribuir a la comprensión de procesos psicológicos, sociales y políticos dentro del programa y las influencias externas sobre el programa.

Las tres razones más frecuentes por las que se realiza una evaluación son 1) juzgar la validez del programa; 2) ayudar a las personas que toman decisiones; y 3) desempeñar una función política (Talmage, 1982). En cada informe concreto, estos propósitos recibirán un énfasis diferente.

Investigación evaluativa definida

La **evaluación** requiere un diseño formal y unos procedimientos para recopilar y analizar los datos de forma sistemática, y así determinar la validez de una práctica educativa específica o prevista. Decir que una práctica o programa tiene **méritos** implica examinar el programa y juzgar su valor respecto a un conjunto de criterios que se aplican de manera relativa o absoluta. Los valores de una práctica educativa pueden ser intrínsecos a la práctica (por ejemplo, los programas de lectura valoran la comprensión lectora), o al lugar (por ejemplo, la cultura y los valores de la comunidad). La práctica educativa hace referencia a un programa, un currículum, una política o una regulación administrativa, una estructura de organización o un producto. Sin embargo, la mayor parte de los ejemplos se extraerán a partir de la evaluación de currículos o de programas.

Un currículum o programa puede implantarse en un lugar o en varios dentro de una sola unidad gubernativa o administrativa local como un sistema escolar de un distrito o una ciudad. El número de lugares en los que la práctica educativa se lleva a cabo, influye en la pregunta de evaluación y en las decisiones sobre el diseño. Los enfoques de evaluación son aplicables a programas y políticas implantadas en uno o varios lugares¹.

Evaluación formativa y sumativa

La distinción entre evaluación formativa y sumativa se ha admitido ampliamente. En la **evaluación formativa**, los investigadores recopilan datos para mejorar un currículum en fase de desarrollo. Las preguntas típicas son: «¿qué partes del programa están funcionando?» y «¿qué necesitamos cambiar y cómo?». Los resultados de la evaluación pueden conducir a la decisión de revisar el currículum, ampliar el estudio de campo y, así obtener más datos, o interrumpir su desarrollo de manera que no se malgasten recursos en un programa que finalmente puede resultar ineficaz.

La evaluación sumativa puede llevarse a cabo una vez que el programa está completamente desarrollado, es decir, cuando funciona correctamente o cumple lo que el programa pretende con pocos efectos perjudiciales. La **evaluación sumativa** determina la efectividad de un programa, especialmente en comparación con los programas con los que compite. Una pregunta típica puede ser: «¿qué programa consigue dichos objetivos de manera más efectiva y eficiente?». La evaluación sumativa puede ayudar a los educadores que buscan apoyo o toman decisiones relacionadas con los nuevos programas, productos o procedimientos.

¹ Nosotros tratamos, brevemente, los estudios a gran escala de múltiples emplazamientos como un método de análisis político.

La tabla 14.1 resume las distinciones entre investigación evaluativa formativa y sumativa.

No debemos olvidar que las audiencias y los empleos de la evaluación formativa y sumativa son muy diferentes. En la evaluación formativa, la audiencia es el personal del programa; en la evaluación sumativa, la audiencia son los usuarios potenciales, agencias y otros funcionarios educativos, al igual que el personal del programa. Tanto la evaluación formativa como la sumativa son esenciales porque se necesitan decisiones durante las etapas de desarrollo de un programa para mejorarlo y, también, cuando está establecido, para juzgar su validez final o determinar su futuro. Desgraciadamente, muchas agencias educativas sólo realizan evaluaciones sumativas.

Dos factores importantes que influyen en la utilidad de la evaluación formativa son el *control* y el *cálculo de tiempo*. La evaluación formativa debe recopilar datos acerca de variables en las cuales los administradores del programa tienen algún control. La información que llega demasiado tarde a los administradores para su aplicación a la mejora del programa resulta, obviamente, inútil.

Las diferencias entre la evaluación formativa y la sumativa ayudan en la selección de un diseño de evaluación. Los estudios formativos se llevan a cabo normalmente por evaluadores internos; los estudios sumativos, por evaluadores externos. Un evaluador externo, sin embargo, puede contratarse para realizar tanto evaluaciones formativas como sumativas para conseguir mayor credibilidad, otro punto de vista tomado del exterior y una actitud neutral hacia la validez de la práctica.

TABLA 14.1: DIFERENCIAS ENTRE LA EVALUACIÓN FORMATIVA Y LA SUMATIVA

	Evaluación formativa	Evaluación sumativa
Propósito	Mejorar el programa	Certificar la utilidad del programa
Audiencia	Administradores del programa y personal	Usuarios potenciales y agencia patrocinadora
Quién debería hacerlo	Evaluador interno	Evaluador externo
Recogida de datos	A menudo multimétodo, informal	Instrumentos válidos / fiables
Muestra	Resolución y/ o probabilidad	Probabilidad
Preguntas formuladas	¿Qué está trabajando? ¿Qué necesita revisarse? ¿Cómo se puede mejorar?	¿Qué resultados se obtienen? ¿En qué situaciones? ¿Qué costes, materiales y formación se requieren?

Fuente: adaptado a partir de Worthen, B. R., Sanders, J. R. y Fitzpatrick, J. L. (1997), *Program evaluation: Alternative approaches and practical guidelines*. (2ª ed.) New York: Longman, Inc. (p. 17).

Criterios para juzgar la calidad de la investigación evaluativa

La segunda edición de *The program evaluation standards* (1994) indica la continua profesionalización y especialización de este tipo de investigación. El *Joint Committee on Standards for Educational Evaluation* representa a importantes asociaciones nacionales de administradores escolares, investigación educativa, evaluación, medición, profesores, orientadores, supervisores, directores de escuela primaria, directores de escuela secundaria, consejos escolares, funcionarios de jefaturas escolares estatales, etc. Las treinta normas fueron desarrolladas por evaluadores profesionales para proporcionar un lenguaje común, un marco conceptual para la evaluación y una base para la autorregulación.

El *Joint Committee* desarrolló cuatro criterios que debería cumplir cualquier estudio de evaluación: utilidad, viabilidad, honradez y precisión. Cada criterio se describe más detalladamente con normas específicas. Aparecen resumidas a continuación².

1. Las *normas de utilidad* garantizan que la evaluación satisfaga las necesidades de información de una determinada audiencia. Estas ocho normas son la identificación de la audiencia, la credibilidad del evaluador, el alcance y selección de la información, la interpretación evaluativa, la claridad del informe, la difusión del informe, la oportunidad temporal del informe y el impacto de la evaluación.
2. Las *normas de viabilidad* garantizan que la evaluación sea realista, económica y oportuna. Estas tres normas son los procedimientos prácticos, la viabilidad política y la efectividad de los costes.
3. Las *normas de honradez* garantizan que la evaluación se realice legal y éticamente, sin olvidar el bienestar de las personas implicadas en la evaluación y los afectados por sus descubrimientos. Estas normas son la obligación formal, el conflicto de intereses, las revelaciones sinceras y completas, el derecho del público a saber, los derechos de las personas, las relaciones personales, la información equilibrada y la responsabilidad fiscal.
4. Las *normas de precisión* garantizan que la evaluación será planteada y que contendrá la información técnicamente adecuada sobre las características de la práctica estudiada que determine su validez. Estas once normas son la identificación de objetos, el análisis del contexto, la descripción de propósitos y de procedimientos, las fuentes de información justificables, la medición válida y fiable, el control sistemático de datos, el análisis de información cuantitativa, el análisis de información cualitativa, las conclusiones justificadas y el informe objetivo.

² Ver Joint Committee on Standards for Educational Evaluation (1994), *The Program evaluation standards* (2ª ed.), Thousand Oaks, CA: Sage.

Las normas no son un libro de recetas con los pasos que hay que seguir sino más bien una recopilación de características comúnmente aceptadas de una correcta práctica evaluativa. En una situación formal específica de evaluación, las elecciones y eliminaciones relacionadas con cada norma están dentro de la competencia del evaluador. Y no sólo eso, sino que además, las normas sirven como guía para los evaluadores, los funcionarios que encargan estudios y las personas que utilizan los informes de evaluación.

ENFOQUES SELECCIONADOS PARA LA EVALUACIÓN

Muchos tipos de estudios reciben el nombre de investigación evaluativa. Los estudios de evaluación responden a necesidades reconocidas por educadores y diversos grupos políticos. Un amplio conjunto de elementos que intervienen en la educación son evaluados. Entre ellos se incluyen:

1. Materiales del currículum (libros de texto, películas, programas software, programas de televisión educativos, vídeos).
2. Programas (*Head Star*, programa sobre lenguajes artísticos, programas ingeniosos, programas preventivos contra el abandono, educación a distancia).
3. Métodos de enseñanza (charlas, lecturas, centros de aprendizaje, descubrimientos).
4. Educadores (directores, profesores, profesores en prácticas, tutores voluntarios).
5. Estudiantes (estudiantes con dificultades de aprendizaje, estudiantes de primaria, de instituto, avanzados).
6. Organizaciones (escuelas alternativas, institutos, escuelas técnico-vocacionales, educación superior).
7. Gestión, utilización de los recursos y costes.

La decisión sobre la entidad que se evaluará es crucial para una evaluación: un grupo, un producto, un método, una organización o un sistema de gestión. La delimitación exacta de la entidad y de todos los elementos que la constituyen ayudan al evaluador a decidir qué componentes son los más importantes para la evaluación.

Un **enfoque evaluativo** es una estrategia para centrar las actividades de la evaluación y para obtener un informe útil. La evaluación es multifacética y puede llevarse a cabo en diferentes fases del desarrollo del programa. Un número de evaluadores han publicado diferentes esquemas, agrupando los enfoques. Cada enfoque cuenta con teóricos importantes, bases lógicas explícitas, discusiones escritas, un grupo de profesionales, estudios de evaluación actuales y críticas. Los enfoques de evaluación más importantes están clasificados en la siguiente lista.

1. *Enfoques orientados hacia los objetivos*, que se centran en la especificación de propósitos y objetivos y determinan hasta qué punto han sido conseguidos.

2. *Enfoques orientados hacia el consumidor*, en los que el tema central es desarrollar información evaluativa sobre «productos» educativos, definidos ampliamente, para que los usuarios de la educación puedan elegir entre ofertas de currículos, de productos educativos, etc.
3. *Enfoques orientados hacia los conocimientos técnicos*, que dependen principalmente de la aplicación directa de conocimientos técnicos profesionales para juzgar la calidad de experimentos educativos, especialmente los recursos y los métodos³.
4. *Enfoques orientados a la decisión*, en los que se subraya la descripción y valoración de un proceso de cambio educativo y los resultados finales, para proporcionar información al encargado de tomar una decisión.
5. *Enfoques orientados hacia el adversario*, en los que el centro de la evaluación es la oposición planificada, desde distintos puntos de vista, de varios evaluadores (pros y contras).
6. *Enfoques naturalistas y orientados hacia el participante*, en los que la investigación naturalista y la implicación de los participantes son esenciales a la hora de determinar los valores, los criterios, las necesidades y los datos para la evaluación.

En este capítulo hablaremos de los tres enfoques que más se utilizan. Los enfoques ilustran la diversidad de la investigación evaluativa: la evaluación orientada hacia los objetivos, la evaluación orientada hacia la toma de decisiones y la evaluación naturalista/participante. Estos enfoques contienen diversas orientaciones para realizar la evaluación, especialmente en los fundamentos de las cuestiones de evaluación. El fundamento de las cuestiones en un estudio basado en objetivos son los objetivos educativos o de los programas. El fundamento de las cuestiones en un estudio orientado a la decisión es la persona encargada de tomar las decisiones, mientras que en la evaluación naturalista el fundamento de la cuestión son las diferentes audiencias (estos es, los *stakeholders*, incluidos los participantes, que resultan afectados por la práctica).

La evaluación orientada hacia los objetivos

La **evaluación orientada hacia los objetivos** determina el grado en el que el grupo objeto de estudio ha logrado alcanzar los objetivos de una práctica. En otras palabras, la evaluación mide los resultados de la práctica. La discrepancia entre los objetivos manifestados y los resultados es la medida del éxito de la práctica. La práctica puede ser un currículum, un curso, la suspensión de un programa en una escuela, la educación de los padres, etc. El **grupo objeto**, el grupo cuyo comportamiento se espera que cambie, pueden ser estudiantes, padres, profesores u otros. Mostraremos los pasos (ver la tabla 14.2) con una evaluación de

³ Son ejemplos.

currículum. Aunque el enfoque parezca simple, el lector debería ser consciente de las decisiones de la investigación que afectan la utilidad de los resultados.

Selección de objetivos mensurables. Un estudio de evaluación mide los objetivos, no las metas abstractas, de la práctica. Las metas en los currículos son, normalmente, declaraciones amplias y generales, que representan valores de la sociedad. Los objetivos son declaraciones concretas que están relacionadas lógicamente con las metas y que son alcanzables a través de la educación. Una meta típica en los estudios sociales es que los estudiantes se conviertan en ciudadanos cumplidores de la ley. El concepto de «ciudadano cumplidor de la ley» alude a un adulto informado que paga impuestos, vota en las elecciones, si se da el caso forma parte de jurados y participa en la democracia. Este comportamiento tiene lugar fuera del sistema educativo y es difícil dejar constancia de ello en un currículum. Sin embargo, los estudiantes pueden ser evaluados según su conocimiento de los deberes y derechos del ciudadano, su participación en elecciones simuladas y sus habilidades en el análisis de temas actuales.

Sólo se evalúan los resultados de los estudiantes definidos como **objetivos de comportamiento**. El *objetivo de comportamiento* es sinónimo de rendimiento o del objetivo medido. Los objetivos de comportamiento son tanto los comportamientos finales como los resultados de los estudiantes (un informe de investigación, una figura de arcilla, una presentación oral, etc.), pero no el proceso que desemboca en el comportamiento final. Los criterios para la consecución del objetivo pueden aparecer o no en la declaración de objetivos. Aquí presentamos cuatro ejemplos de objetivos de comportamiento que presentan niveles diferentes:

Un estudiante, si se le pide, será capaz de deletrear y escribir en mayúscula su nombre correctamente.

Un estudiante será capaz de calcular correctamente el resultado de cualquier problema de divisiones elegido al azar entre los ejercicios de revisión.

Un estudiante realizará un dibujo que será considerado creativo por tres jueces basándose en los criterios de originalidad, flexibilidad y elaboración desarrollados por los jueces.

Al menos el noventa por ciento de los estudiantes será capaz de superar un setenta por ciento en un examen de matemáticas.

El último ejemplo es un objetivo de rendimiento que define el rendimiento mínimo del grupo al 90% y el rendimiento mínimo individual del estudiante. Un análisis del alcance y de la importancia del contenido del currículum indicará qué objetivos son los más importantes.

TABLA 14.2: PASOS EN LA EVALUACIÓN BASADA EN OBJETIVOS

Paso 1	Paso 2	Paso 3	Paso 4	Paso 5
Selección de los objetivos mensurables	Selección del/de los instrumento(s)	Selección del diseño de evaluación	Recogida y análisis de datos	Interpretación de los resultados

Si, además de los objetivos de comportamiento, hay otros objetivos, el evaluador tiene tres opciones: reformular los objetivos en términos de comportamiento sin cambiar el propósito, ignorar los objetivos que no tienen nada que ver con el comportamiento, o comunicar al cliente el hecho de que estos objetivos no se analizarán, pero que podrían ser descritos y valorados con otros procedimientos.

Selección de instrumentos y diseño. Los instrumentos pueden ser exámenes, cuestionarios, autoinformes, escalas de clasificación, sistemas de observación y guiones de entrevista. El instrumento típico es un test estandarizado basado en referencias de logro. Frecuentemente, los evaluadores usan los datos de un programa de examen rutinario. Los datos de un test deben ser válidos para la evaluación. La validez del contenido puede determinarse por un conjunto de expertos locales que comparen el contenido del currículum con los ítems del test. La validez y la fiabilidad de un subtest pueden ser considerablemente menores que para el test entero y debería revisarse en los *Libros anuales de mediciones mentales*. Otras consideraciones son la conveniencia de las normas para el grupo objeto y el tipo de información que se busca. La mayoría de los tests estandarizados basados en normas proporcionan sólo información general sobre los estudiantes comparados con los del grupo normativo.

Los instrumentos basados en criterios pueden usarse para calcular los resultados de los alumnos. Los instrumentos basados en criterios comparan los resultados con criterios predeterminados, patrones o un campo de comportamiento claramente definido: esto es, los responsables de desarrollar el test establecen los objetivos y luego desarrollan los ítems del test. Los instrumentos basados en criterios evalúan tanto a los grupos de estudiantes como a los estudiantes individualmente. Las puntuaciones pueden oscilar entre resultados bajos y altos porque la referencia es un criterio, no a un grupo normativo cuyas puntuaciones se distribuyen según una curva normal.

Los instrumentos basados en criterios deben cumplir los requisitos de cualquier procedimiento de medida. Si un evaluador o un sistema escolar local plantea el desarrollo de un instrumento basado en criterios, resultan necesarios conocimientos de medición y de desarrollo de instrumentos. El instrumento debería ser válido y fiable para los propósitos de la evaluación, aunque el tipo de validez y fiabilidad puede diferir del asociado a los tests basados en normas⁴. El trabajo de campo es esencial.

El diseño más útil para la evaluación basada en objetivos es un diseño de grupos aleatorios o emparejados; no obstante, puede que no resulta factible. Se emplean entonces los diseños semiexperimentales, como el de grupo pretest-postest, las series temporales, o los diseños contrabalanceados. Puesto que la mayor par-

⁴ For a minimal competency test required for, say, high school graduation, *curricular validity* is important. It seems unfair to withhold a diploma from someone who did not learn a competency because it was not in the curriculum. The problems of obtaining evidence of curricular validity and *instructional validity*, a more restrictive term, are myriad. See W. A. Mehrens and I. J. Lehmann (1987), *Measurement and evaluation in education and psychology* (4th ed.).

te de los programas tienen objetivos tanto cognitivos como afectivos, una evaluación comprensiva mediría los diferentes tipos de objetivos si los instrumentos válidos y fiables estuvieran disponibles o se pudieran desarrollar. Los factores en la selección de un diseño son la naturaleza y el número de los objetivos, los grupos objeto, la validez interna y externa y la unidad de análisis. La unidad de análisis es la unidad de datos independiente más pequeña: un grupo o un individuo.

Un informe de evaluación establece la validez y la fiabilidad de las puntuaciones, que son la única fuente de datos para determinar el logro de los objetivos. Pueden citarse estudios de validez y de fiabilidad previos. El informe también proporciona una descripción del proceso de desarrollo y los resultados del estudio de campo para los instrumentos desarrollados localmente. Se anotan los factores relacionados con la validez interna y externa del diseño con sus posibles efectos en los resultados.

Interpretación de los resultados. La evaluación calcula el porcentaje del grupo objeto que alcanza los objetivos predeterminados o valora qué programa, comparado con otros de objetivos similares, es el más apropiado para conseguir los objetivos. Cuando el evaluador examina más detenidamente los objetivos, a menudo, encuentra que están definidos en diferentes niveles de especificidad y que no todos los objetivos pueden evaluarse. Los medios para la selección de objetivos para la evaluación formal son a veces inconsistentes. Como sólo se valoran los logros finales, la evaluación del método se omite. Los resultados pueden sugerir modificaciones de una práctica pero no proporcionan la información completa necesaria para adoptar la práctica en otros lugares.

La evaluación basada en los objetivos es probablemente el enfoque más empleado por varias razones. La mayoría de los educadores estarían de acuerdo en que el alcance exitoso de los objetivos indica tanto el mérito como la validez de una práctica. Los educadores pueden demostrar responsabilidad y un empleo provechoso de los fondos públicos cuando los objetivos se consiguen. Otra ventaja del enfoque basado en objetivos es su metodología altamente definible. Los procedimientos para realizar este enfoque han sido elaborados con mucho esmero, hecho que ejerce cierto atractivo para un número de evaluadores principiantes que deben llevar a cabo evaluaciones anuales. Ningún otro enfoque tiene una tecnología tan elaborada y unas bases tan científicas. Además, si no se alcanzan algunos o todos los objetivos, esto podría conducir a cuestionar los componentes programáticos y a una revisión más detallada de la práctica.

La evaluación orientada a la decisión

La evaluación orientada a la decisión tiene un alcance más amplio que el enfoque orientado hacia los objetivos e implica una teoría del cambio educativo. En este enfoque, «la evaluación es el método de determinar los tipos de decisiones que se tienen que adoptar; seleccionando, recopilando y analizando la información requere-

rida en la toma de tales decisiones; y comunicando esta información a los responsables de la toma de decisiones» (Alkin, 1969, p. 2). Las alternativas de decisión son identificadas por el evaluador y la persona que toma las decisiones, quien determina el mérito y la validez. Las alternativas de decisión pueden ser decisiones rutinarias de mantenimiento o decisiones añadidas que conducen a un cambio que engloba todo el sistema y que requiere una mayor asignación de recursos. Así pues, los estudios de **evaluación orientada a la decisión** pueden realizarse en cualquier punto del proceso de cambio: evaluación de necesidades, planificación del programa, implantación o evaluación del método y de resultados. En la tabla 14.3 resumimos los distintos tipos de estudios de evaluación con sus subsiguientes decisiones.

TABLA 14.3: EVALUACIÓN ORIENTADA A LA DECISIÓN

Evaluación de necesidades	
Evaluación	Situación actual contrastada con situación deseada: necesidad educativa.
Decisión	Selección del problema.
Planificación del programa y evaluación de costes.	
Evaluación	Tipos de programas que concuerdan con los objetivos derivados de la evaluación de necesidades y posibles estrategias.
Decisión	Plan del programa.
Evaluación de la implantación	
Evaluación	Grado en que el programa se implanta tal como se planificó.
Decisión	Modificación del programa.
Evaluación del proceso	
Evaluación	Alcance de consecución de objetivos y resultados del programa.
Decisión	Modificación y mejora del programa.
Evaluación de resultados	
Evaluación	Validez del programa tal y como se refleja en el proceso y los resultados.
Decisión	Certificación y adopción del programa.

Fuente: basado en Stufflebeam, D. L. y otros (1971), *Educational evaluation and decisión-making*, p. 215-239.

1. *La evaluación de necesidades* compara la situación actual y los valores del sistema educativo con los resultados deseados. La evaluación identifica el *contexto*, proporciona la línea-base de datos sobre los logros en un lugar determinado e identifica las necesidades insatisfechas. Las necesidades pueden formularse por los estudiantes, la comunidad, otros grupos o la

sociedad en su conjunto. Los datos son principalmente históricos, descriptivos o comparativos. La evaluación de necesidades conduce a la selección de un programa para conseguir objetivos específicos.

2. *La planificación del programa y evaluación de costes* examina estrategias alternativas, como la adopción de un programa disponible o el desarrollo de uno nuevo para alcanzar los nuevos objetivos. Los investigadores estudian los programas existentes según su factibilidad, su coste y la facilidad para reproducir componentes para alcanzar los objetivos. También examinan la viabilidad del desarrollo local de un programa. La evaluación de costes proporciona información para decidir cómo usar los recursos para conseguir los objetivos deseados. El propósito es valorar una o más estrategias según incida cada estrategia sobre el empleo de personal, de tiempo y de presupuesto, al igual que su potencial para conseguir los objetivos. La evaluación de la planificación del programa conduce a la selección de un plan, incluyendo los procedimientos, materiales, facilidades, equipamiento, esquema, empleo de personal y de presupuestos para el desarrollo o implantación del programa.
3. *La evaluación de la implantación* valora hasta qué punto un programa se desarrolla o implanta como se planificó, e identifica cualquier defecto en el programa. También proporciona información con la que anticipar cambios necesarios para continuar el desarrollo y la implementación del programa. La evaluación de la ejecución revisa y anota lo que ocurre durante el desarrollo del programa. Esta anotación es útil en el análisis retrospectivo de las decisiones, puntos fuertes y débiles de los planes.
4. *La evaluación del proceso* proporciona información sobre el éxito relativo de los distintos componentes de un programa y el grado en que los objetivos y resultados se han alcanzado. El evaluador recopila datos que desembocarán en una mejora inmediata del programa. La recogida de datos puede requerir procedimientos de pruebas y otros métodos. Este tipo de evaluación puede centrarse, también, en el impacto de un programa sobre otros métodos o programas. La evaluación del proceso da como resultado la modificación del programa.
5. *La evaluación de resultados o productos* valora el grado en el que se han conseguido los objetivos. Los datos obtenidos incluyen la evaluación basada en objetivos y otra información de evaluaciones anteriores. La información previa explica por qué los objetivos se alcanzaron o no, y ayuda a la persona encargada de tomar decisiones a eliminar, modificar, mantener o extender el programa para conseguir un uso más amplio. El valor general del programa viene determinado por la manera en que se producen resultados relacionados con los objetivos seleccionados a partir de la evaluación de necesidades. La evaluación de resultados conduce a las decisiones relacionadas con la garantía y aceptación del programa.

Finalmente, el enfoque de evaluación orientado a la decisión se centra en recopilar información gracias a una variedad de métodos para ayudar en las decisiones adoptadas para el desarrollo del programa y conseguir un empleo más amplio. El cambio educativo es una actividad planificada, y la evaluación es una extensión de esta planificación. Algunas posibles dificultades de su uso residen en los conflictos de valores y la discrepancia en los objetivos de un sistema educativo complejo y entre la organización educativa y sus distritos electorales. El enfoque orientado a la decisión da por sentado que la persona encargada de tomar las decisiones es consciente de que puede haber problemas si se efectúa un cambio educativo y, por tanto, quiere obtener información relacionada con estas realidades. No obstante, la colaboración, más que la cooperación, entre el evaluador y la persona encargada de tomar las decisiones, podría dar como resultado datos sesgados. Además, las decisiones normalmente no se presentan como si se tratara de claras alternativas y, a menudo, cambian totalmente de significado con el tiempo. En otras palabras, con este enfoque, es más difícil de lo que puede parecer en un primer momento, especificar y anticipar las decisiones a las que hay que atender. Si el evaluador trabaja estrechamente con la persona encargada de tomar las decisiones, el impacto del esfuerzo de la evaluación dependerá en igual medida de las habilidades del evaluador como del liderazgo de la persona encargada de tomar las decisiones.

A pesar de estas dificultades, el enfoque orientado a la decisión tiene en cuenta la firmeza educativa y metodológica en la evaluación. La evaluación del programa no se basa en un resultado aislado, sino desde que el contexto, los gastos, los planes, el método y los datos resultantes se recopilan y los puntos fuertes, los débiles y las consecuencias se valoran. El grado de realización del programa se lleva a cabo antes de que los resultados del estudiante se valoren. Se destaca el entorno del participante en la evaluación y en la comunicación de información útil. La evaluación puede afectar a las prácticas educativas informalmente a través del proceso de evaluación o formalmente a través de la evaluación final del estudio. El enfoque es flexible, puede aplicarse para objetivos formativos que guíen la toma de decisión a lo largo de todo un proceso de cambio educativo, o puede aplicarse para un objetivo total que demuestre, con un informe, la responsabilidad de decisiones importantes y las bases de dichas decisiones y un informe del sistema actual, logros y decisiones recicladas.

La evaluación naturalista/ participante

Desde 1967, un número de teóricos y usuarios de la evaluación han reaccionado ante lo que consideraban el dominio de los enfoques mecanicistas e insensibles de evaluación en educación. Se expresaron varias preocupaciones: (1) a menudo, los instrumentos técnicamente sofisticados y los informes desviaban la atención de lo que realmente estaba ocurriendo en la educación; (2) muchas evaluaciones a gran escala se llevaron a cabo sin evaluadores, incluso, cuando se debían visitar las clases, y (3) las recomendaciones del informe no reflejaban una

comprensión de los fenómenos detrás de los números, los gráficos y las tablas. Los educadores consideraban que el elemento humano, que estaba presente en la complejidad de la realidad del día a día y las diferentes perspectivas de los que se encuentran en el mundo de la educación, no se reflejaban en la mayoría de los estudios de evaluación. De aquí que estos enfoques reciban el nombre de *naturalistas y orientados hacia el participante* por parte de algunos autores.

La **evaluación naturalista/ participante** es un enfoque holístico que utiliza multitud de datos para proporcionar una comprensión de los valores divergentes de una práctica a partir de las perspectivas de los participantes. La bibliografía y los estudios actuales de evaluación ilustran algunas características de la evaluación naturalista y orientada hacia el participante:

1. Emplea un *enfoque holístico*, que analiza la educación como un esfuerzo humano complejo.
2. Acomoda y protege la *pluralidad de valores* cuando presenta o resume las preferencias dispares sobre la práctica evaluada.
3. Da cuenta de los «retratos» –como han acabado llamándose– de una persona, una clase, una escuela, un distrito, un proyecto o un programa que se sitúa en el contexto más amplio en el que funciona.
4. Depende del *razonamiento* inductivo, que surge de la observación y el descubrimiento de las bases.
5. Utiliza una *multiplicidad de datos* provenientes de distintas fuentes, normalmente dentro de una metodología cualitativa o combinando datos cualitativos y cuantitativos.
6. Emplea un *diseño emergente* para dar un sentido a una práctica específica con sus influencias contextuales, variaciones del método e historias personales.
7. Registra *realidades múltiples* más que una única realidad.

Las estrategias cualitativas se han adoptado para los propósitos de la evaluación⁵. Robert Stake anotó que muchos estudios de evaluación no se usan porque el informe es irrelevante, es decir, que no responde a las necesidades del cliente. Según Stake, «una evaluación educativa es una evaluación respondiente si está orientada más directamente hacia las actividades del programa que hacia los propósitos del programa; responde a las demandas de información de los destinatarios; y si las diferentes perspectivas de valor que considera son mencionadas al dar cuenta del éxito o fracaso de los programas» (1975, p. 14). La **evaluación respondiente** es una antigua alternativa basada en lo que la gente hace natural-

⁵ One example of methodological adoption is the use of focus groups to explore individuals' perceptions about reactions to a particular program, service, or product. The group can be internal staff or a broad range of external stakeholders and consumer groups. The strategy can be employed at the beginning of a program to assess needs, at the mid-point to assess progress and difficulties, or at the end of a program. See Buttram, J. L. (1990). Focus groups: A starting point for needs assessment. *Evaluation Practice*, 11 (3), 207–212.

mente cuando evalúa cosas: observa y reacciona. «La evaluación respondiente acerca los intentos para proporcionar los procedimientos naturales por los que la gente asimila información y llega a la *comprensión*» (Stake, 1973, p.3). El diseño de la evaluación surge a partir de los problemas y asuntos expresados.

Acontecimientos importantes: estrategias informales. La evaluación respondiente es cíclica, incluye acontecimientos que se repiten. Para Stake, «cualquier acontecimiento puede seguir a cualquier acontecimiento, muchos tienen lugar simultáneamente y el evaluador vuelve a cada acontecimiento en muchas ocasiones antes de que la evaluación se termine» (1975, p. 18). En la figura 14.1 los acontecimientos importantes se presentan en forma de reloj, subrayando la naturaleza cíclica del planteamiento.

Los acontecimientos, en la evaluación respondiente, se denominan fases de la investigación. En la fase 1 (desde el medio día hasta las 4, en la figura 14.1) el evaluador habla con los clientes, el personal del programa y las **audiencias**, cualquiera

FIGURA 14.1: ACONTECIMIENTOS IMPORTANTES EN LA EVALUACIÓN NATURALISTA Y ORIENTADA HACIA EL PARTICIPANTE



relacionado directa o indirectamente con el programa, para conseguir una visión general de las diferentes perspectivas y valores del programa. El evaluador observa el programa en funcionamiento. Gracias a estas actividades, el evaluador descubre el significado de los objetivos del programa y conceptualiza los problemas. En la fase 2 (desde las 4 hasta las 7, en la figura 14.1) el evaluador averigua qué datos se necesitan y selecciona los métodos para su recogida. Aunque Stake espera que los observadores y jueces sean el método principal de la recogida de datos, los instrumentos pueden ser apropiados. Los datos se organizan en antecedentes, negociaciones y resultados, incluyendo tanto los resultados deseados como los no perseguidos. En la fase 3 (desde las 8 hasta las 11, en la figura 14.1) el evaluador se preocupa por comunicar los descubrimientos de forma natural. Los retratos pueden incluirse en el informe de investigación convencional, pero normalmente adoptarán la forma de estudios de caso descriptivos, enseres, mesas redondas, artículos periódicos, gráficos o cintas de vídeo, dependiendo de los destinatarios. Sólo las preocupaciones principales de cada destinatario le son manifestadas a los demás. Por último, el evaluador une los informes formales y los otros tipos de informe.

Rol del evaluador. Hay dos aspectos de la evaluación respondente que diferencian este enfoque de las orientaciones principales de las que hemos hablado: el rol del evaluador y la retroalimentación continua. El evaluador responde a los problemas de los destinatarios conforme van cambiando a lo largo del desarrollo del programa y estimula ideas probando las intuiciones y descubrimientos basadas en datos sobre los indagados. La negociación y la interacción forman parte del método para asegurar la exactitud y la comunicación. La comunicación es un proceso de dos etapas en el que los descubrimientos se prueban entre distintas audiencias. Esto puede conducir a que el evaluador vuelva al campo en busca de datos adicionales o altere la manera en que los descubrimientos se redactan para conseguir una comunicación más efectiva. Los resultados presentados en el informe final no deberían sorprender a ninguna audiencia, porque su contenido ha sido discutido ampliamente antes de comunicarlo.

En resumen, el evaluador naturalista reconoce que las preocupaciones de los distintos destinatarios sobre un programa representan distintos valores y necesidades de datos. El evaluador debe seleccionar las preocupaciones que, dentro de los límites de tiempo y recursos, son importantes, legítimas y relevantes. El evaluador es el que realiza el descubrimiento de los valores plurales que enmarcan un programa, independiente de la percepción de cualquier persona aislada encargada de tomar decisiones. Puede utilizarse una variedad de diseños. La flexibilidad de la evaluación respondente asegura que la evaluación será útil para las audiencias.

La evaluación naturalista/participante es normalmente una investigación de estudio de caso, y al igual que con otros enfoques semisubjetivos, existe alguna dificultad en establecer su credibilidad. Aunque la mayoría de los estudios de caso se distinguen más en cuanto al énfasis que en cuanto a la veracidad o la falsedad, los diferentes evaluadores subrayan acontecimientos distintos. La consistencia

metodológica continúa siendo un problema con este enfoque tal y como sucede con la representación de diversos intereses. Algunos piensan que los evaluadores deberían equilibrar los intereses según su propio sentido de la justicia; otros evaluadores adoptan una postura desinteresada y neutral, ofreciendo descripciones y análisis pero sin recomendaciones. Además, la redacción de retratos o los estudios de caso requieren habilidades, preparación y capacidad para manejar datos confidenciales. A pesar de estas dificultades, un estudio de caso bien construido es una evaluación poderosa, con el potencial de ser coherente, prometedora para personas con distintos puntos de vista sobre situaciones diversas, y segura, especialmente sobre los trabajos internos de un programa.

FORMAS DE COMBINACIÓN DE MÉTODOS

Los evaluadores pueden tener preguntas y situaciones en las que resulte útil incluir tanto los métodos cuantitativos como los cualitativos en una evaluación. Estos evaluadores tienden a ser más pragmáticos a la hora de diseñar los estudios de evaluación y menos rígidos para adherirse a un modo de investigación. Obviamente, algunas preguntas de evaluación son respondidas de forma más apropiada a través de los métodos cuantitativos y otras son más adecuadas para las estrategias cualitativas. Si sólo se formulara una pregunta de evaluación en el estudio, ésta requeriría el diseño de un método único⁶.

En un diseño comprensivo en el que sobre una práctica se formula un número de preguntas, puede resultar necesario recurrir a una combinación de métodos. Un estudio de este tipo combina las técnicas cualitativas y las cuantitativas y/o el análisis de datos dentro de diferentes fases del proceso de investigación (Tashakkori y Teddlie, 1998). Puede haber una o múltiples aplicaciones dentro de una evaluación. Cuando se combinan los métodos cuantitativos y los cualitativos en un estudio, el objetivo, el comienzo de los diferentes métodos, y el momento para el análisis de datos de cada método deberían planificarse cuidadosamente por adelantado. Además, el alcance (nacional, estatal, local), la organización temporal para la realización del estudio (normalmente superior a 5 años), y la viabilidad financiera influyen en la selección del tipo y forma de la combinación de métodos. De cualquier otra manera, utilizar métodos combinados puede producir resultados ininterpretables.

Tipos y objetivos de la combinación de métodos

Cinco tipos distintos de combinación de métodos y sus objetivos fueron identificados en un análisis de 57 estudios de evaluación con métodos combinados (Green, y otros, 1989). Puesto que dos de estos cinco tipos se llevan a cabo ex-

⁶ Hemos comentado estrategias multimétodo y de triangulación en los capítulos 10 y 12. La triangulación remite a los multimétodos para recrementar la validez dentro de un modo de investigación.

cepcionalmente, nos concentraremos en los otros tres tipos y hablaremos sobre las formas ilustrativas (tabla 14.4). Hasta ahora hemos hablado de los modos de interrogar cuantitativos y cualitativos como enfoques separados; normalmente, el tipo de muestreo, la técnica de recopilación de datos y el análisis de éstos pertenecen al mismo modo. En los estudios de métodos combinados, la recopilación de datos, el muestreo, la instrumentación y el análisis de datos son decisiones independientes. Los datos cualitativos, por ejemplo, pueden analizarse semiestadísticamente⁷ o el muestreo aleatorio puede aplicarse a la selección de participantes para entrevistas en profundidad o de lugares para la observación de los participantes.

TABLA 14.4: MÉTODOS MIXTOS

Tipo	Objetivo	Formas ilustrativas
Complementario	Elaborar, mejorar, ilustrar, clarificar los resultados de un método con los de otro método.	Formas simultáneas
Relativo al desarrollo	Usar los resultados de un método para desarrollar o informar del muestreo o las técnicas para el método segundo.	Formas secuenciales
Expansión	Extender el campo y el alcance de los resultados gracias al uso de diferentes métodos para diferentes componentes o preguntas del programa.	Formas secuenciales o paralelas

Fuente: adaptado de Green y otros (1989) y Tashakkori y Teddlie (1998).

Los tres tipos de métodos combinados son el complementario, el relativo al desarrollo y el de expansión. Las formas de los *métodos combinados complementarios* utilizan los resultados de un método para elaborar, mejorar, ilustrar o aclarar los resultados de otros métodos. Los datos cualitativos pueden clarificar los resultados de un test estandarizado, o una escala sobre el desarrollo de cada niño puede esclarecer los resultados de un estudio de observación del participante en una guardería. Las formas de los *métodos combinados relativos al desarrollo* aplican los resultados de un método para desarrollar o informar el muestreo o la técnica de recopilación de datos para un segundo método. Por ejemplo, las entrevistas exploratorias pueden usarse para desarrollar items para un cuestionario. Las formas de los *métodos combinados de expansión* extienden el campo y el alcance de los resultados gracias al empleo de diferentes métodos para diversos componentes del programa o para múltiples preguntas sobre la

⁷ Ver tabla 12.2 del capítulo 12.

práctica. Un método de observación participante describe la implementación del programa seguida de una prueba estandarizada que mida los resultados de éste.

Formas de combinación de métodos

Algunas formas de combinación de métodos son usos *simultáneos* de los dos métodos en el mismo componente del programa, *secuenciales* y *paralelos* de ambos métodos, es decir, emplean los dos métodos pero atendiendo a diferentes componentes del programa o preguntas sobre la práctica. A continuación mostramos algunos ejemplos⁸.

Ejemplo 1: Muestreo aleatorio, técnica cualitativa y análisis de contenido. Los participantes, identificados por puntuaciones en una prueba de autoestima, son asignados al azar a los grupos de tratamiento y de control. Las entrevistas en profundidad se llevan a cabo con todos los participantes de ambos grupos antes de que el programa empiece y al final del programa. El análisis⁹ de contenido se realiza separadamente en los grupos de tratamiento y de control; se comparan y contrastan las pautas de cada grupo. (Forma de combinación de métodos: relativa al desarrollo y secuencial).

Ejemplo 2: Muestreo aleatorio, técnica cualitativa y análisis estadístico. Los estudiantes de riesgo son asignados al azar en los grupos de tratamiento y de control. Las entrevistas, relacionadas con los objetivos del programa, se llevan a cabo con todos los participantes de ambos grupos antes e inmediatamente después del final del programa. Las transcripciones de las entrevistas se entregan a un grupo de jueces independientes que puntúan cada entrevista sobre una escala de ocho puntos en diferentes dimensiones de resultado como la probabilidad de éxito escolar, el compromiso con la educación, la autoestima, la aplicación en el trabajo académico y las demostraciones de buenos hábitos nutricionales y de salud. La Estadística inferencial se emplea luego para comparar los dos grupos. Las clasificaciones se relacionan también estadísticamente con las características previas de los estudiantes. (Forma de combinación de métodos: relativa al desarrollo y secuencial).

Ejemplo 3: Muestreo intencional, técnicas cualitativas y cuantitativas, análisis estadístico y de contenido. Los estudiantes se seleccionan para el programa según los requisitos previos del curso. Un observador registra el tiempo invertido en los ejercicios en los que intervienen en cada sesión. El segundo evaluador recurre a la observación participante para escribir pequeñas descripciones de lo que acontece en cada sesión. Dos lectores independientes leen las tablas de frecuencias y las descripciones y anotan lo que en su opinión fueron sucesos importantes. El informe fi-

⁸ Los ejemplos se han adaptado de Patton (1990); Wothen, Sanders y Fitzpatrick (1997) y Tashakkori y Teddlie (1998).

⁹ Utilizamos el término «análisis de contenido» como término genérico para un análisis inductivo de los datos cualitativos en la investigación evaluativa.

nal incluía ambas descripciones completas de los observadores y los lectores. (Forma de combinación de métodos: complementaria y simultánea).

Ejemplo 4: Muestreo intencional, técnicas cualitativas y cuantitativas, análisis estadístico. Los estudiantes se seleccionan para el programa según la necesidad educativa. El evaluador no posee categorías o variables predeterminadas. El evaluador observa las actividades y acontecimientos importantes en el programa, buscando los tipos de comportamiento e intervenciones que surgen. Para cada tipo nuevo de comportamiento o intervención, el evaluador crea una categoría y luego emplea un diseño de muestreo según el tiempo y el espacio para calcular la frecuencia con la que tienen lugar estas categorías de comportamiento e intervenciones. A continuación, la frecuencia de los comportamientos e intervenciones se relaciona estadísticamente con características como el tamaño del grupo, la duración de la actividad, la ratio profesor/alumno y la densidad social/física. (Forma de combinación de métodos: complementaria y secuencial).

Ejemplo 5: Muestreo intencional seguido de muestreo aleatorio, técnicas cualitativas y cuantitativas, análisis de contenido seguido de análisis estadístico. Se utiliza un muestreo intencionado y aleatorio para crear una muestra de hombres y mujeres de tres categorías de edad. Los datos cualitativos se componen de entrevistas semiestructuradas; los datos cuantitativos consisten en un autocuestionario. El análisis cualitativo genera cuatro temas. El análisis estadístico relaciona los autocuestionarios con temas de la entrevista según las características de la edad y del género. (Forma de combinación de métodos: de expansión y secuencial).

Otras formas de diseños de métodos combinados son el empleo multinivel de dos métodos. La investigación multinivel es común tanto en los enfoques cualitativos como en los cuantitativos. Los datos estadísticos recopilados a partir de los alumnos pueden relacionarse con las cualidades de los profesores y las características de la escuela. Los métodos combinados multinivel constituyen otra manera de emplear ambas técnicas. Otras formas de diseños de métodos combinados son las empleadas en la investigación estatal, nacional e internacional a gran escala. Normalmente hay múltiples aplicaciones en cada etapa del estudio: muestreo, recopilación de datos y análisis de datos¹⁰.

Cuestiones de los estudios de métodos combinados

Los estudios de métodos combinados plantean algunas cuestiones. La elección de un procedimiento de muestreo (aleatorio o intencionado) estructura la evaluación y puede suponer la importancia de un método sobre otro. En un diseño de métodos combinados secuenciales sobre los mismos participantes, los investigadores descubrieron que la secuencia de los dos métodos influía sobre el resultado, especialmente cuando el cuestionario se administraba antes de la en-

¹⁰ Ver capítulo 8 en Tashakkori y Teddlie (1998).

trevista (Floyd, 1993). No obstante, se identificaron nuevas cuestiones que no habrían resultado evidentes de haber utilizado cualquier otro método.

Las cuestiones más recientes de los métodos combinados se centran en la metodología, prestando menos atención a la formulación del problema y al establecimiento de conclusiones. Sin embargo, unos cuantos autores han hablado de combinar los dos enfoques en una mayor cantidad de fases de la investigación (Patton, 1990; Creswell, 1995; y Sechrest y Sidani, 1995).

Quizás, el problema del que más se habla sea la síntesis de tipos de datos diferentes. Por ejemplo, la síntesis de respuestas abiertas en un cuestionario con el análisis estadístico de las respuestas cerradas es razonable porque el método dominante es la investigación deductiva. La mayoría de los investigadores sostienen que sintetizar los resultados de diferentes métodos es problemático porque tienen objetivos distintos: la investigación cualitativa sirve para explorar; la cuantitativa para confirmar (Sechrest y Figueredo, 1993). Los dos grupos de resultados aparecen en el informe pero no pueden sintetizarse en muchos estudios.

ANÁLISIS POLÍTICO

El análisis político evalúa las políticas de gobierno para proporcionar a los responsables de elaborar la política recomendaciones pragmáticas orientadas hacia la acción. La política es lo que se pretende cumplir a través de la acción gubernamental y el esfuerzo acumulado de acciones, suposiciones y decisiones de la gente que gestiona la administración pública. Las metas de la administración pública son la igualdad, la libertad, la equidad y la eficiencia, que se consideran parte de la herencia cultural, histórica y legislativa de Estados Unidos. Es prácticamente imposible conseguir estas metas en su más alto grado porque la búsqueda exclusiva de una restringe a las demás. El equilibrio entre estas metas en cualquier momento es el resultado de compromisos entre el sistema político y económico.

Las políticas nacionales y, especialmente, las estatales tienen cada vez una mayor influencia en la política escolar local. Los administradores y profesores de las escuelas crean política, en sentido estricto, conforme realizan sus trabajos día a día. Dos características del sistema educativo permiten la implantación de políticas. En primer lugar, la política es a menudo una suma de ideas vagas y dispares, por lo que se pueden realizar interpretaciones diferentes. En segundo lugar, incluso si una política es coherente y muy específica, los sistemas gubernamentales a través de los que pasa hasta llegar a las escuelas están fragmentados con diferentes advertencias y múltiples prioridades. El gobierno escolar crea oportunidades para que los administradores y los profesores seleccionen los procedimientos de implementación y, de este modo, modificar una política.

Los análisis políticos se centran en (1) la formulación política, especialmente en la decisión de qué problemas educativos tratar; (2) la implementación de pro-

gramas para realizar políticas; (3) la revisión política, y (4) la evaluación de la eficacia y/o eficiencia política. Un programa puede analizarse separadamente de una política o puede definirse como un medio específico adoptado para llevar a cabo una política.

Características del análisis político

Un concepto central en el análisis político es la elección. La *elección* implica la valoración de los *costes de oportunidad* –el curso de acción más valorado se evalúa frente a los *intercambios*–. Los intercambios son acciones alternativas que se abandonan cuando se realiza una elección. Para determinar qué acción escoger, los costes de oportunidad y los efectos, normalmente, se definen en términos económicos o políticos.

Dos enfoques distintos empleados en el análisis político son (1) una aproximación macro, que se basa en modelos económicos como el análisis coste-beneficio y la utilización de bases de datos a gran escala, y (2) una aproximación micro, que se basa en el campo para «obtener realidades» y subraya los métodos cualitativos. Los enfoques macro y micro se distinguen de las suposiciones sobre la elaboración de la política, los valores políticos y las guías para realizar elecciones políticas. Véase la tabla 14.5 para un resumen de las diferencias.

Muchos estudios políticos son eclécticos, combinando métodos cualitativos y cuantitativos. Encontramos raras veces una dependencia total de las pruebas estadísticas en el análisis de las grandes políticas. A pesar de las variaciones en los estudios políticos, hay algunas características de la investigación que la diferencian de otros tipos de investigación. Un número de estas características se comparte con la investigación de evaluación. Generalmente, el análisis político tiende a (1) ser multidimensional; (2) usar para la investigación orientaciones

TABLA 14.5: APROXIMACIONES MACRO Y MICRO AL ANÁLISIS POLÍTICO

	Aproximaciones macro (Bases de datos a gran escala)	Aproximaciones micro (Estudios de campo)
Presunciones de la elaboración política	Científico. Racional.	Decisiones añadidas. Altamente político. Implementación inconexa.
Valores políticos	Excelencia. Eficacia. Utilitario.	Justicia social. Bien común. Descentralización.
Elecciones políticas guiadas por	Modelos abstractos. Árboles de decisión. Ecuaciones matemáticas.	Realidades diversas de la toma de decisiones y la implementación.

deductivas e inductivas; (3) incorporar el futuro al igual que el pasado; (4) responder a los usuarios del estudio, y (5) incorporar valores de forma explícita.

El análisis político, similar a la investigación de evaluación, se centra en variables abiertas a la influencia y la intervención y es sensible a los usuarios. Los usuarios pueden ser numerosos, tener distintas expectativas, valores, presunciones y necesidades. Los valores de los usuarios intervienen en los procesos de definición del problema educativo; en la formulación de las preguntas de investigación, diseño y alternativas políticas, y en el desarrollo de recomendaciones. Los valores educativos siempre forman parte de un contexto cultural. Estos valores, con frecuencia, varían desde el nivel regional hasta el nivel estatal. Además, los valores normativos de la sociedad en general se tienen en cuenta.

Métodos de análisis político

El análisis político incorpora una variedad de metodologías al análisis técnico de problemas políticos. Éstas incluyen una síntesis centrada, análisis secundarios, experimentos de campo, programas experimentales o semiexperimentales a gran escala, programas de prueba a gran escala, entrevistas cualitativas, encuestas e investigaciones de estudio de caso. Las metodologías de las que hablamos en los capítulos precedentes y otras se encuentran definidas brevemente aquí en el contexto del análisis político; los estudios de análisis de costes se describen más detalladamente.

La *síntesis centrada* es la revisión selectiva de materiales escritos y de investigación relevantes para la cuestión política. La síntesis difiere de la revisión de la bibliografía tradicional por el tratamiento de la información obtenida a partir de una variedad de fuentes más allá de los artículos publicados –entrevistas con expertos e interesados, destinatarios, anécdotas, experiencias personales del investigador, documentos sin publicar, notas administrativas y materiales publicados–. Un estudio completo de análisis político puede emplear este método.

El *análisis secundario* es el análisis y reanálisis de las bases de datos existentes. No obstante, las preguntas políticas o modelos de decisión que guían el reanálisis difieren de la pregunta de investigación tradicional en un estudio de metaanálisis. Más que examinar las bases de datos para determinar el estado del conocimiento sobre la dimensión del efecto de una sola práctica educativa, el análisis político genera diferentes modelos de política y preguntas desde las que examinar las bases de datos.

Los *experimentos de campo* y semiexperimentos investigan el efecto o cambio como un resultado de la implementación política. Puesto que los enfoques experimentales tratan de explicar las condiciones educativas existentes, los resultados pueden no ser útiles en el futuro. Las condiciones políticas pueden ser tan dinámicas que los resultados se limitan a ese período concreto de la implementación.

Las *evaluaciones experimentales o semiexperimentales a gran escala* de programas sociales o de programas de servicios sociales en las áreas de la sanidad, la educación, la salud mental o el bienestar público pueden prestar atención a varias preguntas sobre el programa y los lugares. El análisis político semiexperimental a gran escala de los proyectos, normalmente financiado para un período de tres a cinco años, recibe el nombre de evaluación de escenarios múltiples. Algunas razones para llevar a cabo evaluaciones de escenarios múltiples son:

- Determinar el efecto global del programa, cuando se agregan los efectos en todos los lugares.
- Evaluar el programa en una muestra de lugares representativos para estimar el efecto del programa en todos los lugares.
- Determinar si el programa funciona bajo una diversidad de condiciones de implantación.
- Estudiar cómo interactúa el programa con las características específicas de los lugares.
- Comparar la ejecución del programa en todos los lugares para identificar las maneras más efectivas e inefectivas de funcionamiento del programa.
- Facilitar que se compartan las «prácticas eficaces» y otras en todos los lugares.

Como estos lugares son un número de diferentes unidades administrativas locales, se hace un esfuerzo para realizar un muestreo cuidadoso y para estandarizar la implementación del programa y los procedimientos para recopilar y analizar datos¹¹. Los objetivos de las evaluaciones de escenarios múltiples en los que hay que alcanzar un mayor grado de acuerdo se relacionan con aumentar la generalizabilidad de los descubrimientos, ampliar el tamaño de las muestras y, así, aumentar el poder estadístico, y responder a una variedad de asuntos políticos y sociales.

Las *entrevistas cualitativas* pueden ser grupos de discusión con individuos que poseen información decisiva. Los grupos de discusión reaccionan ante temas preespecificados para generar ideas y explorar alternativas posibles o factores causales. Los «estudios prospectivos» que ayudan a los elaboradores de la política a prever las implicaciones y consecuencias de las leyes políticas propuestas, incluyen entrevistas a expertos en un campo para solicitar las más recientes y mejores orientaciones sobre una propuesta y, en algunos casos, utilizar los informes gubernamentales existentes para identificar tendencias.

Las *encuestas* en forma de cuestionario o entrevista, con muestreo aleatorio o intencionado, proporcionan datos sobre las condiciones educativas actuales de grupos o situaciones seleccionadas.

¹¹ Ver Greensberg, Meyer y Wiseman (1995) y Worthern, Sanders y Fitzpatrick (1997) para el tratamiento de planteamientos de diseño.

Los *programas de ensayos a gran escala* a nivel estatal, nacional e internacional sobre la educación son políticas dirigidas por los gobiernos. Estos datos, recopilados a partir de múltiples lugares, pueden determinar el efecto global de un programa o política cuando se agregan los resultados. Un ejemplo es el informe a la nación del estado de la educación estadounidense basado en el programa de Evaluación Nacional del Progreso Educativo. También se realizan estudios similares internacionalmente como el Tercer Estudio Internacional de Matemáticas y Ciencia.

El *análisis de estudios de caso* se emplea frecuentemente para la investigación política porque puede diseñarse para obtener un análisis más global de una situación. Los estudios de caso proporcionan una comprensión más completa de una situación compleja, identifican las consecuencias no perseguidas y examinan el proceso de implantación política, que es útil para futuras elecciones políticas. Por ejemplo, en cinco estudios de caso sobre la implantación de una política estatal de reforma de las matemáticas, dos profesores «no percibieron» el mensaje de la reforma matemática; dos profesores modificaron su manera de enseñar, pero también reformularon la nueva política con las ideas anteriores, y un profesor no reaccionó en absoluto ante la nueva política. «Desde ese momento, los profesores a los que observamos realizaron algunas mezclas importantes entre la instrucción matemática nueva y vieja» (Cohen y Ball, 1990, p. 253). Cada estudio de caso proporcionó diferentes pruebas sobre la necesidad de servicios de apoyo a la implantación de la política.

Análisis de costes

Los elaboradores de la política se preocupan por la eficacia (¿produce la política los resultados deseados?) y los rendimientos (¿se obtienen dichos resultados con el menor coste posible?). «Sólo las alternativas que proporcionen los mejores resultados para un nivel determinado de empleo de recursos o que supongan el mínimo coste para un nivel determinado de resultados nos permitirían aumentar la eficacia de los programas» (Levin, 1981, p. 17). La relación entre coste y eficacia determina si los programas más eficaces son también los más eficientes en lo que a costes se refiere. Por ejemplo, un estudio supuso un gasto de 143 dólares (en dólares de 1985) al año para proporcionar a un estudiante instrucción informática asistida (CAI; *computer-assited instruction*) durante 10 minutos al día usando un sistema de microordenadores con tiempo compartido. De entre todas las alternativas de enseñanza consideradas, CAI en matemáticas fue una alternativa más conveniente, en lo que coste-efectividad se refiere, que las clases de recuperación realizadas por adultos durante 20 minutos al día, o que el aumento del tiempo de instrucción a 30 minutos al día. Resultaba menos efectivo, en lo que a coste se refiere, que las clases de recuperación realizadas durante 20 minutos al día y tenía aproximadamente la misma eficacia en lo que a coste

se refiere que la reducción del tamaño de las clases de matemáticas (Levin, Glass y Meister, 1987).

Levin (1981) presenta cuatro formas de análisis de costes: coste-beneficio, coste-efectividad, coste-utilidad y coste-viabilidad. Estas formas de análisis se diferencian en características distintivas, puntos fuertes y débiles, tal y como presentamos en la tabla 14.6.

TABLA 14.6: FORMAS DE ANÁLISIS DE COSTES

Tipo de análisis	Características distintivas	Puntos fuertes	Puntos débiles
Coste-beneficio (CB)	Resultados medidos en valor monetario.	Compara las alternativas dentro del servicio. Compara los servicios. Resultados expresados como proporciones internas, beneficios netos o proporciones coste-beneficio. Replicable.	Dificultad de convertir todos los resultados en valores monetarios.
Coste-efectividad (CE)	Resultados medidos en unidades de efectos.	Los logros pueden medirse como cambios psicológicos o físicos. Replicable.	La unidad de efectividad debe ser la misma entre programas con las mismas metas.
Coste-utilidad (CU)	Resultados medidos por juicios subjetivos.	Puede integrar múltiples resultados en un solo valor.	Las medidas son subjetivas. No es replicable.
Coste-viabilidad (CV)	Estimar la posibilidad del coste dentro de las restricciones financieras.	Indica si es posible una mayor consideración de alternativas.	No se ocupa de los resultados de las alternativas.

Fuente: «Cost Analysis» de Henry M. Levin, citado en *New Techniques for Evaluation*, editado por Nick L. Smith (Beverly Hills, California: Sage Publications, 1981). Reimpreso con permiso de Henry M. Levin.

El **análisis coste-beneficio** (CB) evalúa las alternativas de decisión comparando el coste y los beneficios para la sociedad. Tanto los costes como los beneficios se calculan en valor monetario. La alternativa que proporcione el mayor beneficio relativo al coste debería ser seleccionada. La ventaja del CB es que puede hacerse un conjunto de comparaciones entre las alternativas dentro de la educación (programas de instrucción diferentes) y entre diferentes tipos de servicio

(educación, salud y transporte). Por ejemplo, los beneficios de tres programas de alfabetización para adultos pueden expresarse como mejoras en productividad, ganancias y servicios autoadministrados para la sociedad, o pueden valorar directamente los cambios en ganancias, logros ocupacionales y servicios autoadministrados en los tres grupos de tratamiento. Las ventajas de CB en la formulación política son evidentes, aunque todos los beneficios y costes tienen que expresarse en términos pecuniarios. Dada la dificultad de conversión de algunos beneficios educativos, como los resultados afectivos del aprendizaje, en valor monetario, el CB sería inadecuado si estos beneficios fueran importantes. Sin embargo, podrían utilizarse otras formas de análisis de costes.

El **análisis coste-efectividad** (CE) compara los resultados del programa (efectividad) con los costes de programas alternativos cuando los objetivos de diferentes programas son similares y cuando se usan medidas comunes de efectividad. La efectividad podría medirse a través de la superación de pruebas de logro, psicológicas o físicas estandarizadas. Las medidas de resultado no necesitan ser convertidas en valores monetarios.

El **análisis coste-utilidad** (CU) identifica qué alternativas, de acuerdo con juicios profesionales, tienen más probabilidades de conseguir los resultados deseados al menor coste. El análisis coste-utilidad puede hacerse para la selección de una alternativa o el análisis puede combinarse con un análisis coste-efectividad. A menudo, el problema de escoger entre varias alternativas implica más que la simple relación entre efectividad y coste. La decisión puede tomarse según la equidad, el impacto social y los valores de los elaboradores de la política. La medida de la utilidad asigna un valor numérico a la conveniencia que cada resultado posible tiene para cada persona encargada de tomar decisiones. Cuando las medidas de utilidad se combinan con medidas de coste-efectividad, la decisión resultante refleja más acertadamente los deseos de los elaboradores de la política. Por ejemplo, un estudio descubrió que en tercer curso, el coste mensual de la ganancia en la nota sobre la formación global en matemáticas era alrededor de 20 dólares por estudiante para CAI y 33 dólares para una enseñanza convencional. Basándonos en la clasificación de utilidad realizada por cada miembro del equipo escolar del valor del éxito escolar del estudiante, la actitud hacia las matemáticas y la informática, CAI tenía una excelente utilidad en lo que al coste se refiere (Fletcher, Hawley y Piele, 1990).

El **análisis coste-viabilidad** (CV) estima los costes para determinar si la alternativa es o no realista, dado un presupuesto existente. Esto se hace a través de la inspección del presupuesto anual después de haber obtenido una estimación del coste. Si los fondos no se encuentran disponibles, un número de elecciones alternativas se generan reorientando los recursos, obteniendo presupuestos más bajos, buscando otros fondos, recomendando la implantación parcial de una decisión política, etc.

El análisis del coste permite a los elaboradores de la política considerar sistemáticamente el impacto de los costes en las diferentes alternativas para tomar decisiones viables, para estimar la probabilidad de varios resultados deseados en relación con los costes y para comparar el coste-efectividad y el coste-beneficios de programas alternativos. Sin embargo, las evaluaciones CB y CE dejan de proporcionar decisiones políticas automáticas entre las alternativas porque los resultados y barreras no cuantificables no forman parte del análisis, y las pequeñas diferencias entre las proporciones de CB y CE dentro de las alternativas no tienen significado en la toma de decisiones. Ninguna forma de CE ni de CU entre las alternativas tenderá a ser conservadora (Levin, 1988). Es más difícil incorporar resultados múltiples que un resultado sencillo a un análisis. La experiencia en el análisis del coste requiere un dominio formal de las herramientas subyacentes tanto del análisis económico como de la evaluación.

EVALUACIÓN EDUCATIVA Y ANÁLISIS POLÍTICO: BENEFICIOS POTENCIALES Y LIMITACIONES

La investigación de evaluación y el análisis político ofrecen muchos beneficios potenciales a la educación, aunque no son una panacea para todos los problemas de la educación. La educación es una actividad compleja dentro de una sociedad muy amplia e interdependiente ante los cambios sociales, económicos y políticos. En este contexto, la investigación de evaluación y el análisis político aportan una perspectiva racional y empírica al área de las decisiones y a la elaboración de la política educativa.

La mayoría de las evaluaciones y de los estudios políticos tienen como objetivo ser utilizados. Un estudio se utiliza si la investigación está relacionada con una decisión distinta o alumbra a los que adoptan decisiones, sobre la definición del problema o sobre nuevas ideas para acciones alternativas.

El último tipo de empleo de la investigación, el procesamiento psicológico de un estudio, no indica necesariamente las decisiones o prescribe la acción que se debe tomar. En una revisión de 65 estudios sobre el uso de los resultados de evaluación en educación, salud mental y servicios sociales, Cousins y Leithwood (1986) descubrieron que dos factores influían sobre la utilización de la investigación: (a) la implementación de la investigación –la calidad y credibilidad del estudio, la relevancia y calidad de comunicación, los resultados del estudio y la oportunidad–; y (b) la decisión o aplicación de la política –necesidades de información, fuentes de información actuales, actitudes de los usuarios hacia la investigación en general, flexibilidad en el proceso de toma de decisiones, etc. La utilización se mejoró cuando:

- la investigación era apropiada por el enfoque, la complejidad metodológica y la intensidad;
- las decisiones que se debían tomar eran importantes para los usuarios y consideradas, de alguna manera, como apropiadas para los datos formalmente recogidos;
- los descubrimientos eran compatibles con las opiniones y expectativas de los usuarios;
- los usuarios estaban implicados en el proceso de investigación y tenían un compromiso anterior con sus beneficios;
- los usuarios consideraban que los datos de los que se informaba eran importantes para sus problemas; y
- una cantidad mínima de información procedente de otras fuentes entró en conflicto con los resultados de la investigación.

Algunos evaluadores y analistas, con una comprensión realista de cómo se elabora la política, proponen tipos de investigación que presenten más probabilidades de utilización. El conocimiento que maneja diversos criterios para ser válido y que contiene una información más completa como el contexto y la implementación del programa, tiene mayores posibilidades de utilizarse. Además, los estudios sistemáticos a largo plazo tienen más probabilidades de influir en los elaboradores de la política, con especificaciones del ámbito completo de temáticas y resúmenes no-técnicos de los descubrimientos.

Ventajas potenciales. La lista de posibles ventajas aumenta conforme más educadores ganan experiencia con la realización y el uso de evaluaciones y estudios políticos. Los beneficios potenciales mencionados más frecuentemente se describen a continuación.

1. *Planifica e implanta mejoras escolares sobre una base sistemática.* Identificar las necesidades, seleccionar las mejores estrategias a partir de alternativas conocidas, revisar los cambios según se producen y medir el impacto de los cambios minimiza la probabilidad de que haya cambios incorrectos y justifica los gastos. De una manera similar, la evaluación sistemática puede evitar las tendencias, la contrarreacción ante la presión política y la simple falta de esfuerzo para dejar saber al público lo que está ocurriendo en las escuelas o con los nuevos programas. La demostración de qué «funciona» es importante para la justificación del programa. La demostración de lo que no «funciona» permite a las personas encargadas de tomar decisiones y con influencia política reconstruir las alternativas consideradas como soluciones.
2. *Probar varios mitos populares sobre los efectos de la educación en el desarrollo del estudiante.* La experiencia profesional, normalmente, dicta la

mayoría de las prácticas de enseñanza y educativas. La información sistemática y, a menudo, sutil para reemplazar o confirmar observaciones casuales, genera los estudios de evaluación.

3. *Demostrar la responsabilidad profesional valorando la calidad de los programas educativos.* Los educadores buscan continuamente maneras para mejorar la calidad de los programas, y tanto la evaluación como el análisis político pueden desempeñar un papel decisivo.
4. *Reducir la incertidumbre sobre las prácticas educativas cuando la experiencia es limitada.* Los efectos colaterales no previstos o los posibles efectos perjudiciales pueden identificarse antes.
5. *Satisfacer los requerimientos de agencias externas para realizar informes que legitimen las decisiones y mejoren el concepto de lo público.* A través de tomas de decisiones factibles y basadas en datos, la imagen de la escuela puede ser más realista. Conforme aumentan las actividades de evaluación, más grupos y agencias de acreditación esperan, si no demandan, alguna forma de investigación de evaluación.
6. *Llevar a cabo análisis del coste-beneficio de programas y prácticas que requieran grandes gastos.* La contabilidad y el mejor empleo de los escasos recursos pueden demostrarse.
7. *Aconsejar a los que tienen influencia en las áreas de decisión política para permitirles anticipar mejor las cuestiones sobre el programa y la política.* Incluso si la mayor parte de las acciones reflejan decisiones añadidas, una serie de decisiones tomadas con antelación puede conducir a importantes cambios programáticos y políticos.

Posibles limitaciones. A continuación describimos las limitaciones mencionadas más a menudo.

1. *El fracaso de muchos estudios para mejorar las prácticas educativas y la formulación de la política educativa.* Las inadecuaciones de la conceptualización y la realización de muchos estudios explica en parte esta circunstancia. Frecuentemente los estudios se realizan sin comprender los factores que afectan al empleo de la información de la investigación, incluso, cuando se realizan correctamente.
2. *La falta de apreciación de que la investigación es sólo uno de los muchos factores que influyen sobre las políticas, las prácticas y las decisiones educativas.* La evaluación y el análisis político no pueden corregir un problema, pero pueden identificar los puntos fuertes y débiles, destacar las conclusiones, exponer las áreas defectuosas y centrarse en alternativas políticas realistas. Corregir un problema es un paso diferente del empleo de los resultados de la investigación.

CREDIBILIDAD DE LA EVALUACIÓN Y DE LAS PROPUESTAS E INFORMES POLÍTICOS

La evaluación y las propuestas e informes políticos se someten al patrocinador o comité que encargó el estudio. Podría tratarse de un sistema escolar local o una agencia estatal o federal. Es aconsejable que el evaluador interno realice una propuesta escrita para justificar sus actividades, incluso, si no hay gastos adicionales o un contrato independiente. Financiar agencias de nivel estatal, nacional e internacional puede poner en circulación una convocatoria de propuestas que contenga las directrices, la cantidad de dinero repartida y la fecha límite de entrega. Resumiremos de manera breve los criterios generales para evaluar propuestas e informes.

Patrones para las propuestas de investigación evaluativa y política

Las propuestas de investigación evaluativa y política se juzgan por el diseño y otras consideraciones. Usamos el término evaluación para incluir el análisis político. Las siguientes preguntas ilustran los criterios típicos:

1. ¿La práctica o política que se evalúa está claramente identificada?
2. ¿Incluye la descripción de la práctica o de la política los objetivos y los resultados esperados? ¿Se describen breve y comprensiblemente los componentes o el alcance de la práctica o la política?
3. ¿Se especifican los objetivos generales de la evaluación?
4. ¿Se especifican todos los destinatarios importantes del estudio?
5. ¿Son importantes para el público las preguntas de evaluación seleccionadas para juzgar la validez de la práctica educativa o de la política?
6. ¿Se proponen procedimientos de recopilación de datos para cada pregunta de investigación? ¿Permite el diseño que se atiendan las preguntas de evaluación surgidas?
7. ¿Son válidas y fiables para este estudio las técnicas de medida seleccionadas?
8. ¿El diseño exige diferentes tipos de datos procedentes de diferentes fuentes para proporcionar una evaluación completa y exacta?
9. ¿Es viable la recogida y análisis de datos para una retroalimentación de confianza e informes serios?

Otros elementos se consideran cuando se acepta una propuesta de evaluación. Aunque no todos pueden aparecer en el plan, el evaluador y el cliente necesitan

un acuerdo mutuo sobre ellos. Las siguientes preguntas ilustran otras consideraciones.

1. ¿La emisión de los informes previstos se produce en su momento de manera que los destinatarios puedan utilizar mejor la información?
2. ¿Se especifican el método y el proceso de realización del informe (escrito, oral o audiovisual)? ¿El plan de evaluación establece un control editorial y el nombre de la persona que emitirá los informes intermedios y finales? ¿Se recogen todos los informes de la evaluación para varios destinatarios específicos?
3. ¿Se protegen los derechos de las personas participantes?
4. ¿El plan de la evaluación es efectivo en lo que al coste se refiere? ¿Se establecen los procedimientos de responsabilidad financiera?
5. ¿Es políticamente viable el esfuerzo de evaluación para la recopilación de datos y la presentación de un informe equilibrado para todos los destinatarios?

Credibilidad de la evaluación y los informes políticos

Normalmente, un informe de evaluación o político es largo, con varios capítulos. El informe consta de una introducción (objetivo y diseño), los descubrimientos organizados por preguntas de investigación o componentes de la práctica, y un resumen con las recomendaciones. Los criterios para juzgar la idoneidad de un informe subrayan dos aspectos: el objetivo y diseño de la evaluación, y los descubrimientos, conclusiones y recomendaciones. Las preguntas siguientes ilustran criterios típicos.

1. ¿Se enuncian el centro de la evaluación junto con el contexto, los objetivos, la descripción de la práctica o política, los propósitos generales del estudio y los enfoques de evaluación o políticos usados?
2. ¿Aparecen especificadas las preguntas de la investigación, la recopilación de datos y los procedimientos de análisis? ¿Se justifican los procedimientos?
3. ¿Se informa de los resultados de manera equilibrada, con revelaciones completas y sinceras, incluyendo las limitaciones del estudio?
4. ¿Es objetivo el informe hasta el punto de que los descubrimientos están basados en hechos verificados y libres de cualquier distorsión por sentimientos y sesgos personales?
5. ¿Se justifican las conclusiones y recomendaciones con la presentación de suficiente información para determinar si estas conclusiones y recomendaciones están justificadas? ¿Se presentan explicaciones alternativas posibles para los descubrimientos, cuando es oportuno?

RESUMEN

1. La investigación de evaluación requiere un diseño formal y unos procedimientos para determinar la validez de una práctica. La investigación de evaluación se emplea para planificar, mejorar y justificar (o no) las prácticas educativas.
2. La validez de una práctica depende de un juicio sobre su valor de acuerdo con los patrones aplicados de manera relativa o absoluta.
3. La evaluación formativa ayuda a revisar una práctica en un ciclo de desarrollo.
4. La evaluación sumativa, realizada cuando una práctica está establecida, determina la efectividad de una práctica comparada con otras prácticas competidoras.
5. Un estudio de evaluación creíble cumple los estándares de utilidad, viabilidad, honradez y precisión.
6. Los enfoques de evaluación más importantes incluyen los orientados hacia los objetivos, los orientados hacia el consumidor, los orientados hacia los conocimientos técnicos, los orientados a la decisión, los orientados hacia el adversario y los naturalistas y orientados hacia el participante.
7. La evaluación orientada hacia los objetivos se centra en comportamientos finales: el alcance hasta el cual los objetivos mensurables de una práctica son alcanzados por el grupo objeto.
8. La evaluación orientada a la decisión, como la evaluación de necesidades, la evaluación de la planificación del programa, la evaluación de la implementación y la evaluación de proceso y de resultados, proporciona información a las personas encargadas de tomar decisiones durante el programa o los procesos de cambio del sistema.
9. La evaluación naturalista/ participante se basa en los problemas de varios *stakeholders*. La multiplicidad de datos, el razonamiento inductivo y los retratos escritos o una serie de estudios de caso caracterizan este enfoque.
10. Un estudio de métodos combinados incluye técnicas cualitativas y cuantitativas y/o el análisis de datos dentro de diferentes fases del proceso de investigación.
11. Los tres tipos de métodos combinados son el complementario, el relacionado con el desarrollo y el de la expansión.
12. El análisis político evalúa las políticas del gobierno para proporcionar recomendaciones pragmáticas a los elaboradores de la política. Se usan tanto las aproximaciones macro (bases de datos a gran escala) como las aproximaciones micro (estudios de campo).
13. Algunos métodos de análisis político incluyen: una síntesis centrada, experimentos de campo, evaluaciones de escenarios múltiples a gran

escala, entrevistas cualitativas, encuestas, ensayos a gran escala, análisis del coste y estudio de caso.

14. Los beneficios potenciales de la evaluación y el análisis político son las mejoras escolares sistemáticas, el análisis de los costes, la valoración de los efectos educativos en los estudiantes, la estimación de la calidad de la educación, la reducción de la incertidumbre en prácticas innovadoras, la legitimación de las decisiones y la información de influencias políticas para anticiparse mejor al programa y a los resultados políticos.
15. Un evaluador, un cliente, y los usuarios pueden juzgar la adecuación de una propuesta o informe de evaluación usando un catálogo de criterios.

EJERCICIOS DE AUTOEVALUACIÓN

Preguntas

1. La investigación de evaluación...
 - a. comprueba las teorías para desarrollar el conocimiento en las ciencias sociales.
 - b. descubre leyes científicas que son generalizables.
 - c. determina el valor de una práctica específica.
 - d. carece de un diseño y procedimientos para recopilar información.
2. La evaluación formativa, frente a la evaluación sumativa...
 - a. precisa de más tiempo.
 - b. utiliza métodos cuantitativos más a menudo.
 - c. puede mejorar un programa; la evaluación sumativa valora su utilidad.
 - d. está más preocupada de la validez y fiabilidad de la medida.
3. ¿Cuál de las siguientes es una descripción incorrecta de la evaluación orientada hacia los objetivos?
 - a. Determina qué objetivos alcanza el grupo objeto.
 - b. Evalúa sólo objetivos mensurables.
 - c. Requiere instrumentos válidos y fiables.
 - d. Proporciona una evaluación del proceso.
4. Una evaluación orientada a la decisión...
 - a. determina la efectividad de una práctica valorando un resultado.
 - b. puede centrarse en las necesidades, la planificación del programa, la implementación y el proceso/resultado o una combinación.
 - c. se centra sólo en las decisiones para mejorar una práctica.
 - d. es inflexible, es decir, se centra sólo en cambios amplios del sistema.

5. En una evaluación naturalista/ participante, el evaluador...
 - a. recopila datos especificados sólo por la persona encargada de tomar las decisiones.
 - b. emplea sólo un método de recogida de datos.
 - c. prefiere la observación y la entrevista casual con un diseño emergente.
 - d. ignora los problemas de los diferentes *stakeholders*.
6. ¿Cuál de las siguientes es una descripción incorrecta de los estudios de métodos combinados? Los estudios de métodos combinados...
 - a. combinan técnicas y/o análisis de datos cualitativos y cuantitativos.
 - b. pueden utilizarse con propósitos complementarios, relativos al desarrollo o de expansión.
 - c. pueden ordenar los métodos de forma secuencial o paralela.
 - d. usarían muestreo aleatorio, instrumentos de medida y análisis estadístico.
7. El análisis político...
 - I. evalúa políticas gubernamentales para proporcionar recomendaciones pragmáticas.
 - II. es multidimensional en la consideración de un número de factores alternativos.
 - III. incorpora explícitamente valores.
 - IV. puede usar un macroenfoque o un microenfoque.

a. I b. I y II c. I, II y III d. I, II, III y IV
8. Los métodos de análisis político incluyen...
 - a. revisiones rigurosas de la bibliografía.
 - b. evaluaciones de escenarios múltiples a gran escala.
 - c. entrevistas cualitativas.
 - d. análisis del coste.
9. La investigación de evaluación y el análisis político pueden...
 - I. planificar e implementar mejoras escolares con una base sistemática.
 - II. comprobar los mitos populares sobre los efectos de la educación en los estudiantes.
 - III. reducir la incertidumbre sobre las nuevas prácticas educativas.
 - IV. informar a las personas con influencia en las áreas de toma de decisiones y de la política para permitirles anticiparse mejor a las consecuencias del programa y de la política.

a. I b. II y III c. I, II, III y IV d. I y III
10. Un informe de evaluación se juzgaría fiable si...
 - a. fuera exacto, pero la descripción de la práctica fuera ambigua.
 - b. estuviera escrito claramente, pero sesgado por el punto de vista de la persona encargada de tomar decisiones.

- c. presentara conclusiones demostradas, pero las fuentes de información no fueran justificables y los instrumentos no fueran válidos.
- d. presentara de forma completa la información relevante y válida mencionando sus limitaciones.

Problemas de aplicación

Analice las siguientes situaciones de evaluación identificando el problema y sugiriendo procedimientos alternativos.

1. Un inspector de enseñanza quiere una evaluación para comparar un nuevo enfoque de estudio independiente con un enfoque de enseñanza regular en las matemáticas de secundaria. Un acuerdo formal escrito con el personal de evaluación del distrito establece los siguientes puntos:
 - a. La evaluación tiene el objetivo de ayudar a los responsables del departamento de matemáticas a decidir si adoptan el enfoque de estudio independiente a nivel del distrito.
 - b. Los procedimientos son para llevar a cabo una comparación local de dos enfoques, incluyendo al 20 por ciento de los profesores de matemáticas y todos sus alumnos.
 - c. Se valorarán el éxito en matemáticas, la actitud del alumno y el entusiasmo del profesor.
 - d. Los profesores se seleccionarán al azar y se les asignará a uno de los dos enfoques.

El inspector decide más tarde que la evaluación debe proporcionar información para mejorar el nuevo enfoque más que decidir sobre la adopción. Cambia el procedimiento para asignar a los profesores y los alumnos, un cambio que hace que no se seleccionen al azar.

El personal de evaluación, asumiendo que el objetivo y el diseño de la evaluación, una vez convenidos, permanecerán igual, recopila y analiza datos como se planeó en un principio. Los evaluadores descubren que la actitud de los alumnos hacia ambos enfoques es similar, pero el éxito del alumno y el entusiasmo del profesor son significativamente mayores en el enfoque de estudio independiente. El informe juzga que este enfoque es superior y recomienda su adopción.

El inspector queda decepcionado de que el informe no haya ayudado a mejorar el enfoque de estudio independiente. Los responsables del departamento se quejan de que los descubrimientos no son seguros porque muchos de los profesores que se asignaron al enfoque de estudio independiente ya estaban predispuestos antes del comienzo del estudio y los alumnos en las clases del estudio independiente obtenían generalmente mejores notas antes de entrar en el programa.

2. Se asignaron dos miembros de una oficina de evaluación del distrito escolar a la superintendente para ayudar a su personal en el análisis de datos cualitativos sobre un proyecto especial de prevención contra el delito. Las necesidades de los datos evaluativos eran las siguientes: la superintendente quería una descripción de las actividades del proyecto para realizar un informe del progreso para la agencia patrocinadora. La agencia quería saber si los contactos entre los jóvenes del proyecto y varias agencias de refuerzo legal –la policía, los tribunales y los agentes juveniles– habían disminuido durante el primer año.

Los evaluadores pasaron más de un mes leyendo la información que se había recopilado: las anotaciones diarias del personal, grabaciones de audiciones, artículos de periódicos, expedientes escolares de cada joven y documentos oficiales de la policía y los agentes juveniles. Los evaluadores desarrollaron un esquema de clasificación en el que cada entrada podía ser categorizada. También contaron la frecuencia con la que cada joven tenía contacto con una agencia de refuerzo legal durante el primer año del proyecto y el año precedente. Dos cosas resultaron rápidamente evidentes: sólo una pequeña proporción de los datos categorizados no produjo revelaciones sobre la naturaleza de las actividades del proyecto durante el año, y el número de contactos con agencias de refuerzo legal fue casi idéntico en los dos años analizados.

Los evaluadores, al no disponer de tiempo, proporcionaron sólo informes imprecisos sobre las actividades e informaron a la superintendente de que el proyecto había demostrado ser inútil para la disminución de contactos con agencias de refuerzo legal. Cuando la superintendente comunicó estos descubrimientos a la agencia patrocinadora, el apoyo fue retirado y el proyecto terminó.

El personal del proyecto se sorprendió, especialmente porque el número de contactos tenía mucho peso en la decisión. Sus miembros señalaron que durante el año del preproyecto estos contactos eran arrestos y comparecencias ante los tribunales, mientras que durante el primer año del proyecto fueron reuniones de supervisión y orientación con los agentes juveniles. Los agentes juveniles afirmaron que estaban satisfechos con los cambios observados en la actitud y comportamiento de los jóvenes.

3. Se añadieron temas sobre ecología en todos los cursos de un colegio de enseñanza media. La comisión del plan de estudios escolar pidió que el departamento de evaluación del distrito evaluara la efectividad de estos temas. Específicamente, la comisión quería saber si los estudiantes aumentaban su conocimiento sobre temas relacionados con el medio ambiente (conservación de especies en peligro de extinción, conservación de recursos escasos) y disminuían su práctica de tirar papeles al suelo en los patios.

Los evaluadores usaron un diseño pretest-postest con un examen y un cuestionario. El examen consistía en el subtest de ciencias de una prueba de logro nacional estandarizada, que incluía higiene, biología y geología. El cuestionario era un autoinforme con el cual los estudiantes se valoraban a sí mismos y a sus compañeros de clase en ciudadanía escolar (respeto hacia los otros estudiantes, respeto a los profesores y respeto a los bienes escolares). El análisis de los datos mostró que no se daba ningún cambio en las puntuaciones pretest-postest del examen o en el cuestionario sobre ciudadanía. La comisión del plan de estudios estaba decepcionada con la evaluación y señaló que no respondía realmente a sus preguntas.

4. Una comisión del profesorado y el director de una escuela de enseñanza media desarrollaron materiales para actividades de *role-playing* con el objetivo de mejorar la disciplina escolar. Luego, se probaron los materiales en la mitad de las clases de ciencias sociales de la escuela. Después de un año de utilización, el comité de profesores solicitó que el personal de evaluación de la oficina central evaluara los materiales. La comisión quería saber si los materiales necesitaban una revisión y si debían emplearse en todas las clases de ciencias sociales de la escuela.

El evaluador entrevistó a los profesores y alumnos que habían usado los materiales. También hizo una encuesta a todo el profesorado con un cuestionario para conocer si los profesores creían que se había producido algún cambio en la disciplina de la escuela. El evaluador preparó un informe que era muy favorable a los materiales, sugería algunos cambios en éstos y recomendaba su uso en todas las clases de ciencias sociales del año siguiente. El evaluador discutió el informe y los descubrimientos con el director del centro.

Cuando el informe salió a la luz, hubo dos grupos que protestaron: la comisión del profesorado responsable de los materiales y los profesores de ciencias sociales que enseñaban los materiales. La comisión rechazó el informe porque no había aportado ninguna contribución a la evaluación. Los profesores de ciencias sociales consideraron que ellos deberían haber sido los primeros en recibir el informe, puesto que eran ellos los que habían usado los materiales y proporcionado la mayor parte de los datos para los descubrimientos.

Directrices para proyectos de investigación

Escribir un proyecto de investigación puede ser el paso más difícil, aunque estimulante, del proceso de investigación. En esta fase, el proyecto completo se sintetiza de una forma determinada. En un proyecto, los investigadores demuestran que saben lo que están buscando, cómo lo buscarán y reconocerán, y explican por qué la investigación merece la pena. Los proyectos cuantitativos reflejan una aproximación deductiva a la investigación educativa. El formato puede ser un esquema relativamente informal proporcionado por un profesor universitario para satisfacer una exigencia del curso, una tesis o disertación formal presentada a un tribunal o un proyecto con estructura presupuestaria requerido por una fundación o una agencia gubernamental.

En el Apéndice A describimos un formato general de proyecto con unas directrices para proyectos cuantitativos, así como para proyectos cualitativos. Aunque muchos de los elementos de un proyecto son similares en la investigación cuantitativa y cualitativa, hay variaciones metodológicas. También describimos la preparación y la crítica de un proyecto.

PROYECTOS DE INVESTIGACIÓN CUANTITATIVA

Redactar un proyecto de investigación cuantitativa, una vez que se han formalizado las preguntas o hipótesis específicas de investigación, requiere muchas decisiones de diseño. Las decisiones relacionadas con los sujetos y la instrumentación son tan importantes como el enunciado del proyecto.

FORMATO GENERAL

Generalmente los proyectos de investigación cuantitativa siguen el siguiente formato. El esquema general contiene todos los pasos necesarios para la formulación y el planteamiento de un estudio de investigación cuantitativo. Los proyectos cualitativos reflejan una aproximación deductiva a la investigación educativa.

- I. Introducción
 - A. Planteamiento general del proyecto
 - B. Revisión bibliográfica
 - C. Preguntas y/o hipótesis específicas de investigación
 - D. Importancia del estudio propuesto

- II. Diseño y metodología
 - A. Sujetos
 - B. Instrumentación
 - C. Procedimiento
 - D. Análisis de datos y presentación
 - E. Limitaciones del diseño
- III. Referencias
- IV. Apéndices

Directrices para proyectos de investigación cuantitativa

I. Introducción. El *planteamiento general del proyecto* consiste en un enunciado claro y preciso del tema de investigación, que ayuda al lector a reconocer la importancia del asunto a investigar y el área educativa al que se dirige. Por lo tanto, el planteamiento del proyecto está vinculado a la bibliografía relacionada, las preguntas específicas y/o hipótesis de investigación y, por último, a la importancia del estudio propuesto. En la introducción se hace un planteamiento directo y conciso del proyecto, idealmente en el primer párrafo, seguido de una descripción de los antecedentes. Es crucial formular el tema del proyecto en términos inteligibles para alguien que normalmente tiene formación científica, pero que puede estar relativamente desinformado sobre el área del proyecto del que vamos a tratar.

La *revisión bibliográfica* presenta la información sobre el tema del proyecto a partir de discusiones teóricas y la investigación previa, proporcionando de esta manera los antecedentes y la necesidad del estudio. La revisión concluye con una exposición de los datos conocidos acerca del tema y las perspectivas del investigador, como por ejemplo una crítica de los diseños de la investigación previa y la identificación de lagunas en la bibliografía. La revisión bibliográfica, aunque se complete en este momento, se amplía frecuentemente a medida que progresa la investigación.

Se formulan ahora las *preguntas y/o hipótesis de investigación específicas*, algunas veces seguidas por definiciones de palabras clave y variables. El formato de pregunta es apropiado para la investigación mediante encuesta. El formato de pregunta o hipótesis es apropiado para diseños de estudios experimentales y no experimentales. El planteamiento de las preguntas y/o hipótesis específicas debería indicar claramente la naturaleza empírica de la investigación, como por ejemplo, un tipo determinado de diseño de investigación. Las definiciones (preferiblemente definiciones operacionales) de las variables siguen al planteamiento de las preguntas y/o hipótesis específicas de la investigación.

La *importancia potencial del estudio propuesto* indica la importancia del estudio en términos de (1) desarrollo del conocimiento e (2) implicaciones para la

investigación posterior y las prácticas educativas. El investigador reflexiona sobre cómo los resultados del estudio pueden hacer una aportación a la teoría y al conocimiento en el área identificada en el planteamiento general del proyecto. Se definen las implicaciones para la investigación futura. Finalmente, el investigador enuncia provisionalmente las implicaciones potenciales para las prácticas educativas. Las implicaciones para las prácticas educativas no son recomendaciones detalladas para temas específicos sino más bien planteamientos generales para profesionales de la educación.

Algunos investigadores prefieren, y algunos formatos requieren, que la importancia del estudio propuesto se sitúe antes de la descripción del diseño de investigación. La mayoría de los investigadores escriben y rescriben esta parte porque es el principal criterio para conseguir la aprobación del proyecto.

II. Diseño y metodología. El diseño y la metodología en la investigación cuantitativa incluyen los sujetos, la instrumentación y el procedimiento de obtención de datos, los análisis y la presentación de los datos y las limitaciones del diseño.

El investigador identifica el tipo de diseño a utilizar (de encuesta, correlacional, experimental, semiexperimental, etc.). Esto orienta al lector a esperar que en el proyecto se discutan ciertos componentes del diseño.

Los *sujetos* se identifican con la descripción de la población de interés y cómo se extraerá la muestra a partir de esta población. Se define el tamaño de la muestra. Se justifica el procedimiento de muestreo y el tamaño de la muestra.

La mayoría de los proyectos contienen una parte en la que se describe cómo se conseguirá la protección de los derechos de las personas. El procedimiento más simple informa sobre cómo se obtendrá el informe de consentimiento y cómo se establecerá la confidencialidad de los datos personales de los individuos. El formulario para el consentimiento de participación en el estudio se presenta normalmente en un apéndice.

La *instrumentación* para el estudio propuesto identifica el/las prueba(s) que se emplearán y explica el motivo de la selección del instrumento o instrumentos como la definición operacional más apropiada de la variable(s) en la(s) pregunta(s) y/o hipótesis de investigación. Si las pruebas ya están creadas, se proporcionan justificaciones de su fiabilidad y validez para el propósito del estudio. Si las pruebas tienen que ser desarrolladas, entonces se definen los pasos para obtener los datos con validez y fiabilidad adecuadas a este instrumento. Los detalles más técnicos de estos documentos se presentan, a menudo, en un apéndice del proyecto.

El *procedimiento* describe la manera en que se realizará el estudio, de forma que puedan estudiarse las relaciones entre las variables. En la investigación mediante encuesta esto incluye la preparación del cuestionario o el esquema de la entrevista, la preparación del entrevistador o las instrucciones a los que van a

manejar los cuestionarios o tests, etcétera. En la investigación experimental y semiexperimental, el procedimiento puede ser más complejo, –identificación de los grupos, especificación del tratamiento experimental y procedimientos para minimizar las variables de error–. Se indican, también, los procedimientos para la sustitución de los sujetos.

El *análisis y la presentación de los datos* definen las técnicas estadísticas que se utilizarán en el análisis de los datos y especifican cómo serán presentados. El investigador define la prueba estadística para *todas* las preguntas y/o hipótesis de investigación y, si es necesario, la razón fundamental para la elección de la prueba. La razón fundamental puede estar en términos del propósito del estudio, tamaño de la muestra y tipo de escalas empleadas en el instrumento de medida. Una técnica estadística se selecciona en base a su idoneidad para la investigación de la pregunta y/o hipótesis de investigación; no se logra nada con el manejo de una técnica complicada cuando una sencilla sería suficiente.

El investigador define, a menudo, los formatos de la presentación de los datos: los tipos de tablas, las figuras y los gráficos que se usarán para organizar y resumir cada conjunto de datos. A menudo, los formatos seleccionados se relacionan con cada pregunta y/o hipótesis de la investigación. Las directrices para la exposición de números y símbolos estadísticos, tablas y figuras se presentan en *Publication Manual of the American Psychological Association* (APA) (5ª edición).

La sección de *limitaciones del diseño* cita los topes que puede identificar el investigador en ese momento: alcance del estudio, el diseño, y/o la metodología. El investigador reconoce que el estudio propuesto se centra sólo en un aspecto concreto de un tema de investigación más amplio, como por ejemplo, los efectos del «etiquetado» de los niños superdotados de primaria, en vez de los efectos del «etiquetado» de todos los alumnos. La definición de las limitaciones de diseño ilustra el conocimiento del investigador de la amenazas a la validez interna y externa en el diseño propuesto. Para el investigador es mejor reconocer las limitaciones que proclamar que tiene el diseño «perfecto». Las limitaciones metodológicas se relacionan específicamente con la validez y la fiabilidad de la instrumentación propuesta o de la prueba aplicada. Las limitaciones se suavizan con sensatez y no deberían ser tan amplias que redujeran la importancia del estudio. Algunas veces, los investigadores prefieren enunciar hipótesis de investigación establecidas para realizar el estudio en lugar de señalar las limitaciones.

III. Referencias. Las referencias citan las fuentes que ha usado realmente el investigador para desarrollar el proyecto y que se citan en el texto del proyecto. Las referencias son aquéllas que se citan principalmente en la revisión bibliográfica; pero, también se incluyen cualquiera de las citadas en el planteamiento general del proyecto, importancia del estudio y, en algunos casos, en la sección metodológica. Todas las fuentes citadas en el proyecto *deberían* incluirse en las

referencias y todas las entradas mencionadas en las referencias *deberían* aparecer en el proyecto.

Un investigador ético no cita resúmenes y *abstracts* de tesis o artículos de revistas como referencias. Tampoco se cita una parte de un estudio mencionado en otra fuente, pero la fuente que contiene esa cita textual sí y, por lo tanto, se menciona en las referencias. El *Manual de Publicación de la Asociación Americana de Psicología* (5ª edición) proporciona las directrices para realizar citas, citas de referencia y listados de referencia.

IV. Apéndices. Los apéndices proporcionan materiales adicionales para aclarar y economizar la presentación y son esenciales para asignar las referencias en el texto. Estos materiales, si se sitúan en los apéndices, se convierten en opciones disponibles para el lector en el momento en que los necesita, en lugar de distracciones en el curso lógico de un proyecto. En los apéndices pueden estar incluidos elementos como los siguientes:

- a. instrucciones a los sujetos,
- b. formularios para el informe de consentimiento de los sujetos,
- c. cartas de permiso para realizar el estudio en una agencia u organización educativa,
- d. estudios piloto,
- e. copias de instrumentos: cuestionarios, esquemas de entrevista, esquemas de observación,
- f. instrucciones para el entrenamiento de las personas que recogen los datos,
- g. credenciales de expertos, jueces u otro personal especializado que se emplearán en el estudio,
- h. modelos de diagramas de diseño de investigación o análisis estadístico,
- i. esquema general del informe final,
- j. propuesta de esquema temporal para la finalización del estudio.

PROYECTOS DE INVESTIGACIÓN CUALITATIVA

Redactar un proyecto de investigación cualitativa, al igual que la redacción de uno de investigación cuantitativa, puede ser el paso más difícil y, sin embargo, estimulante del proceso de investigación. Normalmente, los proyectos de investigación cualitativa son tentativas y con un final abierto para que permitan un diseño emergente. Los proyectos de investigación cualitativa reflejan una aproximación inductiva de la investigación. Los investigadores y los expertos cualitativos reconocen que el grado de especificidad de un proyecto depende *del grado de trabajo preliminar*. En la investigación etnográfica, el trabajo preliminar puede ser la selección del escenario, de la red o la identificación de casos particulares para la investigación. Sin embargo, un proyecto puede ser lo suficientemente

específico para indicar una investigación sistemática que consiga datos válidos sin ser tan específicos que impidan la reformulación del enfoque inicial y del diseño cuando se recojan los datos.

Formato general

El formato general de un proyecto de investigación cualitativa se parece aproximadamente al de un proyecto de investigación cuantitativa con variaciones metodológicas para un modo interactivo de recogida de datos. El esquema general que se proporciona aquí contiene los elementos de un proyecto etnográfico.

- I. Introducción
 - A. Planteamiento general del proyecto
 - B. Revisión de la bibliografía preliminar
 - C. Problemas previstos
 - D. Importancia del estudio propuesto
- II. Diseño y metodología
 - A. Selección del escenario o del grupo social
 - B. Papel del investigador
 - C. Estrategias de muestreo intencional
 - D. Estrategias de recogida de datos
 - E. Análisis inductivo de datos
 - F. Limitaciones del diseño
- III. Referencias o bibliografía
- IV. Apéndices

El formato general varía un poco para la investigación etnográfica.

DIRECTRICES PARA PROYECTOS DE INVESTIGACIÓN CUALITATIVA

I. Introducción. La introducción consiste en el planteamiento general del tema del proyecto, la revisión bibliográfica, los problemas previstos y la importancia potencial del estudio propuesto. Cada uno de estos puntos se explica brevemente a continuación.

El *planteamiento general del proyecto* consiste en un enunciado claro y breve del tema de investigación que permite al lector reconocer su importancia y el área de la educación en la que se enmarca. Se hace un planteamiento directo del proyecto en la introducción, seguido por una descripción de sus antecedentes. Normalmente el planteamiento general del tema se expresa como «describir y analizar» un suceso en curso o un proceso en una orientación de descubrimiento.

La *revisión de la bibliografía preliminar* presenta los marcos conceptuales iniciales que se utilizan en la definición de cuestiones previsibles y es una necesidad del estudio para la identificación de lagunas en nuestro conocimiento. La revisión bibliográfica hace referencia, frecuentemente, a amplias áreas de pensamiento científico (sociológico, psicológico, antropológico o político) de expertos representativos. La revisión *no* es exhaustiva, pero es una revisión inicial que hace explícitas las redes conceptuales con que el investigador introduce el tema del proyecto para centrar las observaciones y entrevistas iniciales. La revisión justifica, claramente, la necesidad de un estudio descriptivo en profundidad mediante el empleo de un planteamiento cualitativo. Por ejemplo, los estudios previos no examinaron los fenómenos con detalle o como un proceso, la investigación relacionada se llevó a cabo con procedimientos cuantitativos o el número de participantes es demasiado pequeño para cumplir los requisitos estadísticos de un diseño cuantitativo.

Los *problemas previstos* se definen como preguntas de investigación amplias y anticipadas que serán reformuladas. Para definir estos problemas, los investigadores, normalmente, han seleccionado el lugar o los participantes y obtenido un permiso provisional o formal para realizar el estudio. En otras palabras, el investigador ya posee información preliminar.

La *importancia potencial del estudio* describe cómo el estudio puede hacer una aportación al (1) desarrollo del conocimiento y (2) a las implicaciones para posteriores investigaciones y a las prácticas educativas. Generalmente, los estudios cualitativos hacen aportaciones al conocimiento proporcionando descripciones detalladas de un hecho natural que no ha sido completamente descrito en la bibliografía. También pueden desarrollar conceptos o una explicación teórica de lo que se observó. La mayoría de los proyectos con un planteamiento de descubrimiento declaran su intención de sugerir investigación posterior con diseños similares para ampliar los resultados con otros diseños y metodologías para verificarlos.

II. Diseño y metodología. El diseño y la metodología, en un proyecto de investigación cualitativa, incluyen el escenario o el grupo social seleccionados, el papel del investigador, las estrategias de muestreo intencional, las estrategias de recogida de datos, el análisis inductivo de los datos y las limitaciones del diseño.

El investigador identifica el proyecto como un *diseño de estudio de caso*, con el empleo de técnicas cualitativas. Esto orienta al lector a esperar que ciertos componentes de diseño se discutan en el proyecto.

El *escenario seleccionado* se describe en términos que ilustran su idoneidad para investigar los fenómenos y procesos definidos en los problemas previstos. Por ejemplo, si los problemas previstos se enfocan desde los procesos de enseñanza, entonces, el lugar debería referirse a cursos o clases (escenarios sociales) en los que tengan lugar los procesos de instrucción. Resulta esencial una descripción de las características del escenario: colegio público o privado, tipo de

empresa educativa y su papel o función en la sociedad, procesos y actividades típicas, tipos de participantes, etcétera. La descripción ayuda a identificar si el escenario es representativo de otras instituciones y, así, permitir la prolongación de los resultados.

El grupo social seleccionado se describe para justificar que es probable que los miembros del grupo estén informados sobre los problemas previstos. Por ejemplo, si el grupo social está formado por miembros de una asociación profesional o por padres de un grupo de apoyo para niños financiado por un programa, entonces se describen las funciones del grupo y su organización. Debería haber una relación lógica entre la información potencial que se obtendrá a través del contacto personal y los problemas previstos.

El investigador define *el papel de la investigación* a asumir mediante la recogida de datos. En este punto, el investigador puede describir su papel sólo en términos generales, por ejemplo, observador participante o primordialmente entrevistador. Debido a que el papel de la investigación afecta a las relaciones en la recogida de datos interactiva, la función debería quedar delimitada lo más posible en términos de las relaciones sociales esperadas y conjuntos de cometidos durante la recogida de datos. Además, el papel de la investigación debería ser apropiado a los problemas previstos.

En un proyecto se describen las *estrategias de muestreo intencional*. El investigador reconoce que las estrategias se seleccionarán en el campo a medida que progresa la investigación. El propósito de un muestreo intencional es obtener pequeñas muestras de casos ricos en información de individuos, de escenarios sociales o de procesos.

La mayoría de los proyectos define cómo se logra la protección de los derechos de las personas. El procedimiento más simple informa acerca de cómo se obtendrá el informe de consentimiento y como se establecerá la confidencialidad de los datos personales de los individuos. El formato para el consentimiento de participación se encuentra, normalmente, en un apéndice e incluye la garantía del anonimato relacionado con los datos del lugar, la organización o el grupo social y los individuos.

Se definen las *estrategias de recogida de datos*. Aunque las estrategias específicas de recogida de datos surgirán en el campo, la intención de emplear múltiples métodos debería ser explícita para permitir la comprobación de los datos. El investigador especifica qué estrategias utilizará: observación participante, el formato de las entrevistas en profundidad y artefactos que previsiblemente se recogerán en el campo. Algunos investigadores etnográficos especifican ciertas estrategias de recogida de datos para cada tema de investigación en los problemas previstos. El investigador también define la *duración prevista del trabajo de campo*, el límite natural de tiempo del suceso a observar, como por ejemplo, un semestre, el programa completo, etc.

Se definen asimismo los formatos que adoptarán los datos, como por ejemplo, notas de campo, observaciones resumidas, registros de entrevista, transcripciones y elaboraciones. Debido a que los datos recogidos pueden llegar a ser «montones» de información, algunas veces, los investigadores definen cómo catalogarán, guardarán y recuperarán la información ya sea de forma manual o por ordenador. Aunque en un proyecto, las estrategias de recogida de datos se redactan como estrategias planificadas, deben indicar que el investigador es consciente de la necesidad de tomar decisiones en el campo para obtener datos válidos en tanto se reformulen los problemas previstos.

La descripción del *análisis inductivo de datos* incluye las estrategias para facilitar el descubrimiento en el campo mediante el análisis provisional, la codificación de temas y el desarrollo de categorías y técnicas de búsqueda de patrones y el establecimiento de la potencialidad de los patrones. Algunas veces, se proporciona un diagrama del proceso inductivo en un apéndice. Frecuentemente, también, aparecen en un apéndice unas cuantas páginas de una transcripción codificada o de notas de campo. En ocasiones, los investigadores citan los programas de *software* que manejan para el tratamiento de datos.

La sección de *limitaciones del diseño*, al igual que en la investigación cuantitativa, cita las barreras que el investigador puede identificar en ese momento: el alcance del estudio, la metodología y el diseño. La limitación del alcance se relaciona con el planteamiento del proyecto. Los problemas previstos se orientan, normalmente, sobre uno de los aspectos de los posibles centros de investigación en el escenario seleccionado, como por ejemplo, describir procesos de enseñanza pero no la evaluación del profesor o la eficacia de estos procesos para el aprendizaje del alumno. Las limitaciones metodológicas se refieren a las posibles dificultades en la idea del papel de la investigación, el muestreo intencional y los hechos naturales que el etnógrafo no puede interrumpir de forma legítima.

Los proyectos cualitativos también cuentan con limitaciones de diseño: validez, poder de reflexión y generalización de los resultados. Los hallazgos de un diseño de estudio de casos no son generalizables, pero sin un diseño de estudio de caso no pueden lograrse otros propósitos de la investigación. Los investigadores discuten las estrategias que tienen intención de emplear para minimizar las amenazas a la validez y a la extensión de los resultados.

III. Referencias o bibliografía. Las referencias para un estudio cualitativo son similares a las de uno cuantitativo si se emplea el manual de estilo APA. Algunos investigadores prefieren usar el CMS con su formato bibliográfico.

IV. Apéndices. Los apéndices en un proyecto cualitativo proporcionan materiales adicionales para la presentación económica al igual que en los proyectos cuantitativos. Sin embargo, los elementos de los apéndices son los apropiados para el proyecto. En los apéndices de un proyecto etnográfico pueden incluirse elementos como los siguientes:

- a. cartas de permiso que garanticen el acceso al lugar o al grupo social,
- b. acuerdos de información acerca de los derechos humanos con los participantes más significativos,
- c. protocolos para la obtención del informe de consentimiento entre un grupo social,
- d. breves ejemplos hipotéticos de notas de campo y observaciones resumidas, registros de entrevista y transcripciones,
- e. descripción de las personas que nombrarán a participantes para el muestreo de casos de prestigio,
- f. unas pocas páginas de transcripción codificada o notas de campo de un estudio piloto,
- g. lista de registros y artefactos que se sepa que están disponibles en el lugar o a través del grupo social,
- h. propuesta de esquema temporal para la finalización del estudio.

PREPARACIÓN Y CRÍTICA DE UN PROYECTO

La preparación de un proyecto implica dos tareas adicionales: la conformidad con el formato y el estilo requerido para el proyecto, y la obtención de críticas de un proyecto breve antes de que se redacte en su forma final. Los investigadores principiantes a menudo consultan libros sobre redacción de proyectos¹.

Preparación de un proyecto

La mayoría de facultades y universidades han desarrollado su propio manual de estilo y formato o han seleccionado un manual de estilo como guía, por ejemplo, *el Publication Manual of the American Psychological Association* (2001) (5ª edición) [Manual de estilo APA] o *el The Chicago Manual of Style* (14ª edición), revisado y ampliado (1993) [CMS]. El *formato* se refiere al patrón general de organización y ordenación del proyecto. El *estilo* hace referencia a las reglas de ortografía, empleo de mayúsculas, puntuación y tipo de letra a seguir en la preparación del proyecto. Aunque ninguno de estos manuales publicados hace referencia específica a proyectos de investigación, muchas facultades y universidades siguen el formato general y el estilo adoptados por uno de esos manuales autorizados. El estilo y formato de las referencias o bibliografía, las secciones y encabezados y el estilo de redacción son diferentes de un manual a otro. Mientras que el estilo de la APA elimina prácticamente el uso de notas a pie de página, el

¹ Ver, por ejemplo, Lock, L. F., Spirduso, W. W. y Silverman, S. J. (1999). *Proposals that work: A guide for planning dissertations and grant proposals* (4ª ed.).

CMS alienta el uso amplio de notas a pie de página explicativas para citar fuentes específicas de los hechos presentados, las perspectivas metodológicas, los comentarios e información adicional para una mayor comprensión del texto. Además, en el estilo de la APA, la lista de referencias sólo contiene aquéllas que se citan en el texto; el estilo del CMS puede incluir referencias que proporcionan el conocimiento de los antecedentes para el tema o para el investigador.

En un proyecto de investigación se espera que tanto la construcción gramatical como la ortografía y la puntuación sean correctas. Asimismo, se desaconseja el empleo de abreviaturas y contracciones. Los investigadores experimentados saben que «escribir significa volver a escribir y volver a escribir y volver a escribir...». Los investigadores principiantes deberían evitar pronombres personales como, por ejemplo, yo, mío, nosotros y nuestro. Si la primera palabra de una frase es un número, o si el número es nueve o menor, normalmente se expresa como una palabra. De lo contrario, normalmente los números se expresan con guarismos.

El formato también incluye las páginas preliminares. En un proyecto esto se refiere, normalmente, a la portada y, quizá, al índice de contenidos si sobrepasa las quince páginas. La portada contiene el título del proyecto, el nombre del autor y la fecha para un proyecto de un curso. La portada de un proyecto de tesis doctoral puede incluir el título, el nombre del autor, el grado o título al que se aspira, el nombre y la localización de la facultad o universidad que otorga el título, la fecha de presentación o defensa y un espacio para las firmas del tribunal. El título debería ser breve (quince palabras como máximo) y debería describir el propósito del estudio propuesto. El índice de contenidos básicamente es el esquema general del proyecto con el número de página en el que se encuentran los principales apartados, las referencias o bibliografía y los apéndices.

Todos los proyectos se escriben a máquina y, en cuanto a la mecanografía y escritura del proyecto, rigen las mismas orientaciones que para las becas. La versión final debería corregirla el autor cuidadosamente. Los procesadores de texto facilitan la tarea de escritura enormemente. Entre las características comunes se incluyen la numeración automática de página, la posibilidad de editar «en la pantalla» las palabras, frases y párrafos, corrector ortográfico, diseño de página automático y colocación automática de encabezados y pies de página. La corrección de las pruebas es esencial porque los correctores ortográficos *no* corrigen ni la gramática ni el significado de una palabra en una frase.

Críticas de un proyecto

Después de completar el borrador de un proyecto, los autores lo leen de forma crítica desde el punto de vista de los criterios de investigación apropiados para el propósito y el diseño del estudio. Muchos de estos criterios se sugirieron en capítulos anteriores. Además de la autocrítica, los investigadores proporcionan

un borrador a sus colegas para conocer su opinión. A continuación vemos algunas limitaciones habituales de los proyectos.

1. ***El tema del proyecto resulta trivial.*** Aquellos temas que son sólo de interés periférico para los educadores o muestran escasa probabilidad de hacer aportaciones valiosas al conocimiento educativo rara vez se aprueban de forma entusiasta. El tema debería estar relacionado con alguna de las principales corrientes del conocimiento educativo, pensamiento académico, investigación y prácticas educativas.
2. ***El tema no está delimitado.*** Un tema de proyecto debería estar centrado, tanto por razones de investigación como de práctica. Los diseños no pueden conseguir datos válidos para cada posible variable, ni los investigadores cualitativos pueden abarcar cuestiones extremadamente amplias en un estudio simple. Los investigadores experimentados saben cuánto tiempo consumen los procesos de la investigación, desde la conceptualización inicial de una idea hasta el informe final. Los investigadores, de forma racional, pero a regañadientes, delimitan el tema. Las preguntas y/o hipótesis específicas de la investigación o los problemas previstos se definen de tal manera que la delimitación del enfoque es aparente.
3. ***Los objetivos del proyecto son demasiado generales.*** Algunas veces, las hipótesis se definen de una manera tan amplia y con términos tan generales que sólo el diseño de la investigación puede transmitir lo que realmente trata el estudio. Si el diseño de la investigación no se corresponde, de forma lógica, con las preguntas y/o hipótesis específicas de investigación o las cuestiones de investigación cualitativas, entonces el estudio planificado no es capaz de cumplir los objetivos propuestos. No considerar las variables extrañas es un serio error en un proyecto cuantitativo. Aunque, generalmente, las preguntas de investigación se redacten en términos más generales, el punto de vista particular reduce el enfoque.
4. ***La metodología carece del detalle apropiado para el estudio propuesto.*** Los proyectos cuantitativos deberían ser lo suficientemente detallados respecto a los sujetos, instrumentación y análisis de los datos como para permitir su replica. Los proyectos cualitativos, debido a su naturaleza inductiva, son menos específicos en ciertos aspectos. Una investigación cualitativa, sin embargo, debe ser lo suficientemente detallada al informar sobre un *posible* muestreo intencional, las estrategias planificadas de recogida de datos y las técnicas de análisis inductivo de datos. Esta especificación asegura a un comité de evaluación que el investigador es consciente de las decisiones que tomará después. Mucha de la especificidad, tanto de los proyectos cualitativos como de los cuantitativos, depende del grado del trabajo previo del investigador.

5. *No se tratan suficientemente las limitaciones del diseño.* Ningún diseño es «perfecto»; cada diseño representa una serie de elecciones para equilibrar los requisitos del diseño y los patrones más frecuentes con los intereses viables de restricciones fiscales, logísticas y temporales. Los investigadores experimentados son conscientes de sus decisiones y citan las principales limitaciones del diseño con su base lógica y los procedimientos para minimizar sus efectos en el estudio.

Glosario

Abstracciones sintetizadas: generalizaciones y explicaciones resumidas de los principales descubrimientos del estudio, con formatos variados según la disposición cualitativa seleccionada.

Acuerdo: tipo de fiabilidad basada en la consistencia de los índices o en las observaciones entre dos o más personas.

Alpha de Cronbach: tipo de consistencia interna para elementos con respuestas graduadas.

Análisis político: evalúa las políticas del gobierno para proporcionar recomendaciones prácticas a los responsables de tomar decisiones.

Análisis de conceptos: estudio que aclara el significado de un concepto por la descripción de su significado genérico, los diferentes significados y el uso apropiado para el concepto (investigación filosófica).

Análisis de coste-beneficio: evalúa alternativas de decisión por comparación de costes y beneficios para la sociedad.

Análisis de coste-eficacia: compara los resultados de prácticas similares con relación a sus costes, cuando los diferentes programas tienen los mismos objetivos y medidas.

Análisis de coste-utilidad: compara prácticas alternativas para determinar la práctica que es más probable que produzca los resultados más deseados al menor coste.

Análisis de coste-viabilidad: estima el coste de una práctica para determinar si la práctica es una consideración realista dentro de un presupuesto existente.

Análisis de covarianza (ANCOVA): test estadístico inferencial que se emplea para ajustar estadísticamente el efecto de una variable relacionada con la variable dependiente.

Análisis de sendas: procedimiento estadístico que usa correlaciones entre un conjunto de variables que se han ordenado lógicamente para reflejar las relaciones casuales.

Análisis de varianza (ANOVA): procedimiento estadístico inferencial para la determinación del nivel de probabilidad de rechazar las hipótesis nulas con dos o más significados.

- Análisis de varianza factorial:** análisis de procedimiento estadístico de varianza con dos o más variables independientes que permite la comprobación de cada variable independiente y la interacción de las variables.
- Análisis durante el proceso:** análisis de datos cualitativo frecuente y regular para ayudar a tomar decisiones e identificar los temas que surgen.
- Análisis histórico:** colección sistemática y de críticas de documentos que describen e interpretan acontecimientos pasados.
- Análisis inductivo:** categoría y patrones que surgen a partir de los datos en lugar de estar impuestos a los datos antes de su recogida.
- Analogía:** metáfora, hecha explícita con la palabra «al igual» o «como», que compara dos fenómenos diferentes.
- Asignación aleatoria:** procedimiento usado para asignar sujetos a los diferentes grupos, de manera que cada sujeto haya tenido igual oportunidad de ser seleccionado en cada grupo.
- Atenuación:** ocultación de la medida de relaciones entre dos variables a causa de la poca fiabilidad de las pruebas.
- Audiencias (en investigación evaluativa):** personas que se guiarán por la evaluación para tomar decisiones, cualquiera directa o indirectamente conectado con el programa.
- Auditoria:** registro de técnicas de gestión de datos y reglas de decisión que documentan la «cadena de evidencia» o el «sendero de decisión».
- Autenticidad:** fiel reconstrucción de las percepciones múltiples de los participantes.
- Base de datos (en búsquedas por ordenador):** fuentes indexadas por un servicio de referencia particular.
- Bibliografía primaria:** búsqueda original de estudios o escritos efectuada por investigadores y teóricos.
- Bibliografía relacionada:** bibliografía relevante para el problema o está relacionada con el diseño de alguna manera esencial.
- Bibliografía secundaria:** síntesis de la bibliografía primaria: teórica, empírica o ambas.
- Biografía comentada:** estudio en el que la presencia del investigador se conoce en la descripción y por su punto de vista.
- Biografía:** como tipo de investigación, puede incluir biografías individuales, autobiografías, historias verídicas, y la historia oral de una vida.
- Bivariado:** correlación entre dos variables.

- Buscador:** servicio que cataloga y recupera información de Internet.
- Características demandadas o exigidas:** posible fuente de sesgo cuando cualquier aspecto de un estudio revela el propósito de la investigación y los sujetos responden de forma diferente debido a su conocimiento del propósito.
- Caso único:** investigación que se realiza con sujetos individuales para estudiar los cambios en el comportamiento asociado con la intervención o la eliminación del tratamiento.
- Caso:** situación determinada, seleccionada por el investigador, en la que se describe algún fenómeno por medio de los significados de los participantes de los acontecimientos y procesos.
- Categoría:** término abstracto que representa el significado de temas relacionados.
- Categorías emic:** vistas de los miembros, como por ejemplo acciones y explicaciones, que son características del entorno o de la gente.
- Categorías etic:** puntos de vista de la situación de los profanos, como por ejemplo los conceptos y las explicaciones científicas del investigador.
- Chi-cuadrado:** procedimiento estadístico no paramétrico que se utiliza con datos nominales para comprobar relaciones entre la frecuencia de observaciones en las categorías de las variables independientes.
- Ciencia:** generación y la verificación de una teoría a través de la investigación.
- Clasificación de campo:** en investigación cualitativa, la adquisición de datos de las relaciones social, espacial, y temporal en el lugar para conseguir un sentido del contexto total.
- Codificación:** proceso de la división de los datos de acuerdo a un sistema de clasificación.
- Código:** abreviatura para un tema o categoría.
- Coefficiente de contingencia:** también, coeficiente pi. Índice simple que muestra el grado de relación en una tabla de contingencia.
- Coefficiente de regresión múltiple:** indicador de la relación combinada de varias variables con otra variable.
- Coefficiente de correlación:** número que se calcula para indicar el tamaño y la dirección del grado de relación entre dos variables.
- Coefficiente de determinación:** coeficiente de correlación cuadrado que indica el porcentaje de varianza acumulada en una relación.
- Coefficiente de regresión:** factor que se usa en la regresión múltiple para ponderar la contribución de cada variable a la ecuación.
- Coefficiente phi:** ver coeficiente de contingencia.

- Colecciones de documentos:** objetos materiales relacionados con acontecimientos, grupos, personas, u organizaciones actuales o del pasado, que revela procesos sociales, significados y valores.
- Comparación permanente:** técnica cualitativa para identificar las características distintivas entre temas, categorías, o modelos por la anotación de similitudes y diferencias.
- Comparaciones *post hoc*:** pruebas estadísticas que se emplean con pares de medias, se realizan, generalmente antes de la prueba estadística, de todas las medias.
- Comparaciones planificadas:** prueba estadística utilizada para seleccionar parejas de medias.
- Comparativa:** ver investigación comparativa.
- Conclusiones estadísticas:** amenaza a la validez interna de un estudio debido a los usos inapropiados de procedimientos estadísticos.
- Conjunto de datos:** grandes colecciones de datos disponibles de fuentes primarias archivados por propósitos de investigación.
- Consistencia interna:** tipo de fiabilidad de la prueba en la que la homogeneidad de los elementos de un instrumento se evalúa después de una administración del instrumento.
- Constructo infraestimado:** acuerdo que falla en la incorporación de facetas importantes de lo que se está midiendo.
- Constructo:** abstracción compleja que no es observable directamente, como por ejemplo, la ansiedad, inteligencia, autoestima; una combinación de conceptos llena de significados.
- Contaminación:** en investigación cuantitativa, un tipo de observador que sesga los resultados debido al conocimiento que posee acerca del estudio.
- Contexto:** descripción de un lugar y de ambientes seleccionados, escenarios sociales, participantes y duración de la recogida de datos.
- Contraste planificado:** comprobación estadística predeterminada de pares de significados seleccionados.
- Control:** empeño para borrar o de lo contrario tener en cuenta factores o variables además de la variable independiente que podrían afectar a la variable dependiente.
- Correlación:** medida de relación que usa un coeficiente de correlación.
- Credibilidad:** grado en el que los resultados se aproximan a la realidad y se juzgan para que sean razonables y fiables.

- Cristalización:** estilo analítico por el cual el investigador colapsa la segmentación, la categorización y la búsqueda de patrones en un periodo de inmersión extenso, muy intuitivo dentro de los datos.
- Crítica externa:** procedimientos analíticos que se llevan a cabo para determinar la autenticidad de la fuente, esto es, si la fuente es el documento original, un documento falsificado o una variante del documento original.
- Crítica interna:** procedimientos analíticos para determinar la credibilidad de los planteamientos en una fuente; la exactitud y fiabilidad de los hechos.
- Cuestionario:** conjunto de preguntas o enunciados escritos que evalúa actitudes, opiniones, creencias e información bibliográfica.
- Cuestiones de contingencia:** preguntas que al ser contestadas de una cierta manera provocan otras cuestiones.
- Datos:** resultados obtenidos por una investigación a partir de los que se confeccionaran las interpretaciones y conclusiones.
- Definiciones operacionales:** definición de una variable conseguida por la asignación de significado a una variable por la especificación de las actividades u operaciones necesarias para medirla, clasificarla o manipularla.
- Desviación típica:** medida de variabilidad; un índice numérico que indica la dispersión media o la difusión de las puntuaciones alrededor de la media.
- Descripciones cualitativas:** narraciones detalladas de personas, incidentes y procesos.
- Descripciones narrativas:** reseñas detalladas de personas, acontecimientos y procesos.
- Descriptiva:** investigación que describe un fenómeno actual o pasado cuantitativamente.
- Deseabilidad social:** tendencia de los sujetos a contestar a los ítems de forma que parezca deseable a los demás.
- Desgaste del sujeto:** amenaza a la validez interna en la que la pérdida de sujetos afecta a los resultados.
- Diagnóstico a través de la carpeta de trabajos:** diagnóstico alternativo mediante el cual los trabajos del alumnado demuestran su rendimiento y son recogidos, organizados y evaluados.
- Diagnóstico basado en rendimientos:** tipo de test en el que se asegura la competencia de un estudiante mediante la observación del rendimiento en un contexto original y auténtico.

- Diagnóstico:** medición de una variable y el empleo de los resultados, por ejemplo, comprobación de los alumnos y presentación de puntuaciones.
- Diagramas de barras:** representación gráfica de la frecuencia de variables nominales.
- Diagrama de caja:** ilustración gráfica de la variabilidad en un grupo de puntuaciones.
- Diagrama de dispersión:** scatterplot, representación gráfica de la relación que se elabora al presentar, de forma visual, la intersección de las puntuaciones de los sujetos en dos variables.
- Diferencia estadística significativa:** término usado en la evaluación de los resultados de estadísticas inferenciales que las diferencias, muy probablemente, no se deben al cambio.
- Diferencial semántico:** tipo de escala en la que los sujetos dan una respuesta entre pares de adjetivos con relación a un concepto u objeto.
- Difusión del tratamiento:** amenaza a la validez interna por la que los sujetos son conscientes de otras condiciones de la variable independiente.
- Diseño A-B:** diseño de caso único que compara la frecuencia del comportamiento durante la línea base (A) con las condiciones del estudio (B).
- Diseño A-B-A:** diseño de caso único que compara la línea base (A) con el estudio (B) y luego con la línea base (A).
- Diseño de inversión:** en investigación de caso único, se refiere al final de la condición del tratamiento y restaurar las condiciones de la línea base.
- Diseño de investigación:** plan que describe las condiciones y procedimientos para la recogida y el análisis de los datos.
- Diseño de pretest y postest con grupo de control:** diseño experimental puro en el que uno o más grupos de sujetos asignados al azar recibe un pretest, un tratamiento, y un postest, y un grupo de sujetos asignados al azar recibe sólo el pretest y el postest.
- Diseño de pretest y postest con un solo grupo:** diseño preexperimental en el que un grupo de sujetos recibe un pretest, un tratamiento, y un postest.
- Diseño de pretest-postest con grupos no equivalentes:** diseño semiexperimental en el que se comparan grupos que no se han asignado al azar para el tratamiento.
- Diseño de reversión:** también de reserva; en investigación de sujeto singular, también se refiere a la finalización de la condición del tratamiento y la restitución de la condición de línea de base.

- Diseño de series temporales interrumpido de grupo único:** diseño semiexperimental en el que las observaciones múltiples de la variable dependiente se hacen antes y después del tratamiento.
- Diseño de series temporales sin grupo de control:** estudio de series temporales semiexperimental que compara el grupo de tratamiento con un grupo de control.
- Diseño de sólo postest con grupo de control:** diseño experimental puro en el que uno o más grupos de sujetos asignados al azar recibe un tratamiento y un postest, y un grupo de sujetos asignados al azar sólo recibe un postest.
- Diseño de sólo postest con grupos no equivalentes:** diseño preexperimental en el que uno o más grupos de sujetos, no asignados al azar, recibe un tratamiento y un postest, y un grupo de sujetos sólo recibe el postest.
- Diseño de solo postest con un solo grupo:** diseño preexperimental en el que un grupo de sujetos recibe un tratamiento y un postest.
- Diseño de estudio de caso:** fenómeno que selecciona el investigador para entenderlo en detalle, independientemente del número de ambientes, escenarios sociales, o participantes en un estudio.
- Diseño emergente:** plan de investigación en el que cada paso depende de los resultados de los datos de campo previos.
- Diferencia estadística significativa:** término usado en la evaluación de los resultados de estadísticas inferenciales que indica que las diferencias muy probablemente no se deben al cambio.
- Diseño *ex post facto*:** ver investigación *ex post facto*.
- Diseño experimental:** investigación en la que las variables independientes se manipulan para investigar relaciones de causa-efecto entre la variable independiente y la variable dependiente.
- Diseños de línea base múltiple:** Un tipo de diseño de caso único que usa varios sujetos, tipos de comportamiento, o situaciones simultáneamente.
- Diseños de series temporales:** Diseños de investigación en lo que un grupo de sujetos se mide repetidamente antes y después del tratamiento.
- Diseños preexperimentales:** tipo de diseño experimental que carece normalmente de validez interna.
- Distribución de frecuencia:** presentación de un conjunto de puntuaciones por el número de veces que se obtiene cada puntuación.
- Distribución de la muestra:** distribución de frecuencia de las posibles muestras de una población.

Distribución normal: distribución de puntuaciones simétrica con forma de campana con la misma media, mediana y modo.

Distribuciones asimétricas negativas: distribución de puntuaciones que tiene un número desproporcionadamente grande de puntuaciones altas.

Distribuciones asimétricas positivas: distribución de puntuaciones que tienen un número desproporcionadamente alto de puntuaciones bajas.

Documentos: registros de acontecimientos del pasado con forma de cartas, diarios, notas anecdóticas, y documentos conservados, frecuentemente, en colecciones.

Duración de la grabación: tipo de procedimiento de grabación del observador en el que se graba la duración del comportamiento.

Ecuación de predicción de regresión múltiple: procedimiento estadístico para el uso de varias variables para predecir algo.

Efecto *Hawthorne*: tendencia de la gente a actuar de una forma distinta debido al hecho de que son sujetos en un estudio.

Efectos de los sujetos: cambios en el comportamiento del sujeto como respuesta a la pertenencia al estudio.

Efectos del tratamiento: amenaza a la validez interna en la que los sujetos están seguros de otras condiciones de la variable independiente.

Efecto del investigador: también, contaminación del investigador. Una amenaza a la validez interna en la que el tratamiento diferencial de los sujetos por parte del investigador afecta a los resultados.

Efecto del pretest: amenaza a la validez interna en la que la realización de un pretest puede afectar a los resultados.

Elaboración de entrevistas: reflexiones de los entrevistadores sobre su papel y su entendimiento, reacciones de los entrevistados, información adicional, y ampliaciones de los significados de la entrevista.

Empírico: guiada por la evidencia, los datos o las fuentes.

Énfasis holístico (en etnografía): subcasos de datos que están relacionados con el contexto social del fenómeno estudiado.

Enfoque evaluativo: estrategia para centrar las actividades de evaluación y elaborar un informe útil.

Enseres: también artefactos, objetos materiales, relacionados con un hecho pasado o actual, de una persona u organización que revelan procesos sociales, significados y valores.

- Entrevista estándar o típica de respuesta abierta:** tipo de entrevista cualitativa en la que se solicita a los participantes que contesten a las mismas preguntas y en el mismo orden para obtener datos de los significados del participante; ver entrevista en profundidad.
- Entrevistas a informadores decisivos:** entrevista en profundidad de un individuo que tiene un conocimiento, estatus, o niveles de comunicación especiales.
- Entrevista con guía de aproximación:** también, entrevista guiada; los temas se relacionan de antemano, pero el investigador decide el orden y la expresión de las cuestiones durante la realización.
- Entrevistas de casos atípicos:** aplicación especial de entrevistas en detalle de personas consideradas influyentes, prominentes, y bien informadas en una organización o una comunidad.
- Entrevistas en profundidad:** uso de una guía general de entrevista para unos pocos temas y pruebas seleccionados (no un conjunto de preguntas estandarizadas); una conversación de al menos una hora con una meta.
- Entrevistas estructuradas:** entrevistas sistemáticas que emplean las mismas preguntas y categorías de respuesta.
- Entrevistas fenomenológicas:** tipo de entrevista en profundidad específica para examinar los significados o la esencia de una experiencia vivida.
- Entrevistas informales:** preguntas que surgen a partir del contexto inmediato y que se plantean en el curso natural de los hechos; no hay predeterminación de redacción o de temas de pregunta.
- Entrevistas profesionales y de historias de vida:** entrevistas que suscitan la narración de historias descriptivas de la vida de los individuos o historias profesionales.
- Equivalencia:** tipo de fiabilidad de la prueba en el que las puntuaciones de formularios paralelos o del mismo instrumento, entregados al mismo tiempo, se correlacionan.
- ERIC (*Educational Resources Information Center*):** base de datos comprensiva e índice de Bibliografía educativa.
- Error típico:** derivación típica de la distribución de la muestra.
- Error tipo I:** rechazo de una hipótesis nula cuando es verdadera.
- Error tipo II:** fallo en el rechazo de una hipótesis nula cuando es falsa.
- Escala Likert:** tipo de escala en la que los sujetos expresan grados de acuerdo o desacuerdo con una propuesta.

Escala: elementos del cuestionario que consisten en gradaciones, niveles o valores que describen varios niveles de algo.

Escalas de medida: propiedades que describen las relaciones entre números.

Estabilidad: tipo de fiabilidad del test en la que se correlacionan las puntuaciones de un mismo instrumento tomadas en dos ocasiones.

Estadística descriptiva: procedimientos estadísticos que describen algo.

Estadística inferencial: procedimientos que indican las probabilidades asociadas con decir algo sobre las poblaciones basadas en los datos a partir de las muestras.

Estadísticos: procedimientos para la organización y el análisis de datos cuantitativos.

Estrategias: muestreo cualitativo y técnicas de recogida de datos que se refinan continuamente a lo largo del proceso de recogida de datos para incrementar la validez de los datos.

Estrategias interactivas: observación etnográfica o entrevista etnográfica como una estrategia de recogida de datos.

Estudio científico: búsqueda del conocimiento con el uso de procedimientos conocidos en recogida, análisis e interpretación de datos.

Estudio cualitativo: estudio en profundidad con el uso de técnicas cara a cara para la recogida de los datos en su entorno natural.

Estudio disciplinado: investigación dirigida de tal manera que puede examinarse el argumento cuidadosamente.

Estudio fenomenológico: investigación que describe los significados y la «esencia» de una experiencia vivida.

Estudios críticos: investigación cualitativa en la que el investigador se compromete a presentar la manipulación social, las estructuras sociales de cambio opresivo, y pueden tener metas emancipatorias.

Estudios de caso: investigación cualitativa que examina un «sistema específico» o un caso en detalle a través del tiempo, empleando múltiples fuentes de datos que se encuentran en el entorno.

Estudios de métodos combinados: estudio que combina técnicas cualitativas y cuantitativas y/o análisis de datos dentro de las diferentes fases del proceso de investigación.

Estudios de predicción: investigación en la que los comportamientos o niveles se predicen mediante una o varias variables.

Estudios evolutivos: investigación que investiga el cambio de los sujetos a través del tiempo.

Estudios preliminares: búsqueda limitada por el uso de uno o más servicios de referencia, número de años para ser evaluado por expertos, o número de fuentes deseadas; normalmente para seleccionar un problema de investigación.

Etnografía: descripción de una cultura o grupo social o sistema.

Evaluación de escenarios múltiples: investigación cualitativa diseñada para informar de la práctica en cada escenario y hacer generalizaciones a través de los sitios.

Evaluación alternativa: procedimientos para medir el rendimiento mediante preguntas con respuestas elaboradas, distinta de las tradicionales de «lápiz y papel».

Evaluación formativa: evaluación usada para mejorar una práctica o programa en curso.

Evaluación naturalista/participante: evaluación holística que utiliza el multi-método para proporcionar un entendimiento de los valores divergentes de una práctica a partir de las perspectivas de los participantes.

Evaluación orientada a la decisión: evaluación que sustituye información por decisiones específicas, como por ejemplo necesidades de acuerdo, planificación del programa, implementación del programa, o resultados.

Evaluación orientada a los objetivos: evaluación que determina el grado en el que se consiguen los objetivos de una práctica por un grupo objetivo.

Evaluación respondente: evaluación diseñada para suplir información sobre los temas y preocupaciones de los destinatarios; usa un diseño emergente para proporcionar un entendimiento del programa.

Evaluación sumativa: valoración diseñada para determinar el mérito, el valor, o ambos de una práctica desarrollada y hacer recomendaciones con respecto a su adopción y a su uso extendido.

Evaluación: estudio que requiere un diseño formal para coger y analizar los datos para determinar el valor de una práctica; ver investigación evaluativa.

Evidencia basada en contenidos del test: tipo de evidencia de la validez en la que las puntuaciones representan un significado, interpretación, rasgo o teoría esencial.

Evidencia basada en la estructura interna: la validez se evidencia mostrando las correlaciones adecuadas entre los ítems.

Evidencia basada en procesos de respuesta: la validez se evidencia mostrando la consistencia entre las propuestas de respuesta y las respuestas contestadas.

Evidencia basada en relaciones con otras variables: la validez se evidencia mostrando las condiciones adecuadas con otras medidas.

Experimental pura: investigación en la que los sujetos se asignan a los grupos al azar, por lo menos se manipula una variable independiente, y las variables extrañas se controlan para investigar la causa de una o más variables independientes sobre la variable dependiente.

Explicación: teoría o generalización analítica que define las relaciones causa-efecto en los enunciados simples.

Extensión de los descubrimientos/resultados: estudios cualitativos que permiten a otros entender situaciones similares y aplicar los resultados en investigaciones posteriores.

Fiabilidad de la prueba o del test: indicación de la consistencia de la medida.

Fiabilidad: grado en que son consistentes las medidas de un test.

Formato abierto: tipo de ítem de cuestionario en el que los sujetos escriben la respuesta a la pregunta.

Formato cerrado: tipo de ítem de cuestionario en el que los sujetos eligen entre opciones propuestas.

Función de la investigación: relaciones adquiridas por y atribuidas al investigador en la recogida de datos interactiva, apropiadas para el objetivo del estudio.

Fuentes de variabilidad: influencias sistemáticas, erróneas, y extrañas relacionadas con el diseño de investigación.

Fuentes primarias: en investigación analítica, un documento o testimonio de un testigo de un acontecimiento.

Fuentes secundarias: en investigación histórica, los documentos o testimonios de individuos que no observaron o participaron directamente en el acontecimiento.

Generalizabilidad: grado en el que los hallazgos de un estudio pueden usarse como conocimiento sobre otras poblaciones y situaciones.

Generalización: en investigación analítica, la interpretación de los hechos que se enfocan sobre la manera o la razón por la que ocurre un acontecimiento.

Grados de libertad: concepto matemático que indica el número de observaciones que son libres de variar.

Gráfico de barras: representación gráfica de la frecuencia de variables nominales.

Gráfico de caja-y-bigote: ilustración gráfica de la variabilidad de un conjunto de puntuaciones.

Grupo de comparación: uno de los grupos en los que la ejecución del comportamiento se compara en un experimento.

- Grupo de control:** en un experimento, los sujetos que no reciben tratamiento.
- Grupo de tratamiento:** ver grupo experimental.
- Grupo experimental:** sujetos que reciben la condición que el experimentador supone que cambiará de comportamiento.
- Grupo objeto:** grupo cuyo comportamiento se espera que cambie como resultado de una práctica.
- Hechos:** en investigación analítica, descripciones de quién, qué, cuándo, y dónde ocurrió un evento, obtenido a partir de evidencia decisiva.
- Hipótesis alternativas posibles:** posibles explicaciones, además del efecto sobre la variable independiente, para relaciones causa-efecto.
- Hipótesis de investigación:** planteamiento provisional de la relación esperada entre dos o más variables.
- Hipótesis estadística:** hipótesis definida en términos de resultados estadísticos.
- Hipótesis nula:** planteamiento estadístico formal que, normalmente, es un planteamiento de falta de relación entre dos o más variables.
- Histograma:** ilustración gráfica de una distribución de frecuencia, se usan barras para representar la frecuencia de cada puntuación.
- Historia oral:** formato de investigación histórica que graba palabras habladas y testimonios de individuos sobre el pasado.
- Historia:** amenaza a la validez interna en la que los eventos o incidentes que ocurren durante la investigación afectan a los resultados.
- Historiografía:** estudio de los procedimientos que usan los diferentes historiadores en su investigación; también un estudio de las revisiones e interpretaciones cambiantes del pasado.
- Índice:** tipo de escala de medida en la que los números se expresan con un significado total como índices.
- Inferencia baja:** tipo de observación en la que el observador registra lo que sucede con respecto a los comportamientos específicos.
- Informe de consentimiento:** obtención de permiso de los individuos para participar en una investigación antes de que esta comience.
- Infraestimación:** descenso de una medida de una relación entre dos variables debido a la falta de fiabilidad de los instrumentos.
- Instrumentación:** amenaza a la validez interna en la que el cambio de los instrumentos y la falta de fiabilidad afecta a los resultados.
- Interacción:** único efecto de los diferentes niveles de las variables independientes sobre la variable dependiente.

- Interpretación analítica:** en la investigación histórica, una explicación casual para un evento determinado.
- Intervalo:** tipo de escala de medida en la que los números tienen un rango que es igual a los intervalos entre rangos.
- Intervalo de confianza:** margen que describe los valores de población probables.
- Investigación:** proceso sistemático de recogida y análisis de datos con un propósito concreto.
- Investigación acción:** los profesores utilizan los métodos de investigación para estudiar los problemas de la clase.
- Investigación analítica:** análisis de los documentos para investigar conceptos y eventos históricos.
- Investigación aplicada:** investigación que se dirige a un campo de práctica común y se preocupa de la aplicación y el desarrollo del conocimiento basado en la investigación.
- Investigación básica:** investigación que comprueba o perfecciona una teoría; no está diseñada para ser practicada inmediatamente.
- Investigación comparativa:** tipo de investigación cuantitativa no experimental que examina las diferencias entre grupos.
- Investigación correlacional:** investigación en la que se recoge para cada sujeto la información de, por lo menos, dos variables para investigar la relación entre las variables.
- Investigación cualitativa:** investigación que presenta hechos en una narración con palabras; ver también estudio cualitativo.
- Investigación cuantitativa:** investigación que presenta los datos con números.
- Investigación de campo:** investigación que ve el entorno como una situación natural en la que el investigador recoge datos durante un tiempo prolongado.
- Investigación descriptiva:** investigación que describe el estatus actual de algo.
- Investigación evaluativa:** diseñada para asegurar el valor de una determinada práctica en términos de valores operativos en el sitio(s).
- Investigación *ex post facto*:** investigación que investiga los eventos que ya han ocurrido e implican relaciones causa-efecto a partir de los resultados.
- Investigación experimental:** ver diseño experimental.
- Investigación observacional:** investigación de campo en la que los datos obtenidos por observación se recogen a través de procedimientos sin interferencias.

- Investigación por cuestionario:** Investigación en la que la información se recoge mediante las respuestas de los sujetos a preguntas escritas.
- Investigación por encuesta:** evaluación del estado actual o de las opiniones, creencias y actitudes mediante cuestionarios o entrevistas a partir de una población conocida.
- Investigación por test:** investigación en que se usan como datos las puntuaciones del test de los sujetos.
- Kuder-Richardson:** tipo de fiabilidad de consistencia interna para elementos puntuados correctos o incorrectos.
- Línea de base:** La primera frase de una investigación de caso único, en la que se registra el comportamiento antes de hacer ningún cambio.
- Listas de correo:** (servidor de lista de correos). Lista compartida para comunicar a través de correo electrónico sobre temas específicos.
- Longitudinal:** estrategia de investigación en la que los datos cuantitativos sobre los sujetos se recogen a través de un periodo de tiempo.
- Maduración:** amenaza a la validez interna en la que los cambios madurativos en los sujetos como por ejemplo envejecer o estar más cansados, o tener hambre, afecta a los resultados en una investigación cuantitativa.
- Matriz de correlaciones:** tabla que presenta correlaciones entre muchas variables.
- Media:** medida de tendencia central, la media aritmética de los resultados.
- Mediana:** medida de tendencia central, el punto o puntuación de una distribución que está en la mitad.
- Medidas discretas:** también, no reactivas. Uso de métodos para la recogida de información de tal manera que el sujeto ignore ser un participante de la investigación.
- Medidas de tendencia central:** índices resumen de un conjunto de puntuaciones que representan la puntuación típica en una distribución.
- Medidas de variabilidad:** índices numéricos que indican el grado de dispersión de la puntuación a partir de la media.
- Mérito:** alcance de una práctica en relación con las estimaciones, los patrones y las limitaciones prácticas de la adopción de un escenario potencial.
- Metáfora:** comparación implícita entre cosas que esencialmente no tienen que ver unas con otras.
- Método científico:** proceso de investigación secuencial de definición de un problema, definición de una hipótesis, recogida y análisis de datos, e interpretación de resultados.

Métodos de investigación: procedimientos usados para recoger y analizar los datos.

Moda: medida de tendencia central; la puntuación más frecuente.

Modalidad de investigación: también denominada tradición investigadora; colección de prácticas eclécticas de investigación basada en un conjunto de preferencias metodológicas, opiniones filosóficas o ideológicas, cuestiones de investigación y resultados factibles.

Mortalidad experimental: amenaza a la validez interna en la que la pérdida de sujetos afecta a los resultados.

Muestra de tiempo: tipo de registro de observación en el que se emplean periodos concretos de tiempo para observar comportamientos.

Muestra: grupo de sujetos a partir de los que se recogen los datos; a menudo representa una población.

Muestreo aleatorio: procedimiento empleado para asignar sujetos a grupos diferentes por el que cada uno de los miembros posee las mismas posibilidades de ser asignado a cada grupo.

Muestreo estratificado: forma de muestreo al azar en la que primero se divide una población en subgrupos o estratos y se seleccionan los sujetos de cada subgrupo.

Muestreo en cascada (de bola de nieve): estrategia cualitativa en la que cada participante sucesivo se nombra por un grupo o individuo previo.

Muestreo incidental: método sin la probabilidad de selección de sujetos que están disponibles o accesibles.

Muestreo comprensible: tipo de muestreo en la investigación cualitativa en el que se examina cada participante, grupo, ambiente, acontecimiento u otra información.

Muestreo intencionado: estrategia para elegir pequeños grupos de individuos que probablemente tengan conocimiento e información sobre el fenómeno de interés; selección de casos sin necesidad o deseo de generalizar a todos los demás.

Muestreo no probabilístico: procedimiento de muestreo en el que no se conoce la probabilidad de seleccionar elementos a partir de la población.

Muestreo no proporcional: muestreo estratificado en el que el número de sujetos seleccionados a partir de cada estrato no se basa en el porcentaje de la población representada por cada estrato.

Muestreo de variación máxima: en búsqueda cualitativa, una estrategia para probar por anticipado diferentes significados del fenómeno o concepto de investigación.

- Muestreo por conveniencia:** muestreo no probabilística en el se seleccionan sujetos accesibles o aprovechables.
- Muestreo por cuotas:** método de muestreo no probabilístico en el que los sujetos se seleccionan en proporción a las características reflejadas entre la población.
- Muestreo por grupos:** forma de muestreo de probabilidad en el que los sujetos se agrupan en primer lugar de acuerdo a las unidades naturales.
- Muestreo probabilístico:** sujetos confeccionados a partir de una población con probabilidades conocidas.
- Muestreo proporcional:** tipo de muestreo estratificado en el que el número de sujetos seleccionados a partir de cada estrato se basa en el porcentaje de sujetos entre la población de cada estrato.
- Muestreo sistemático:** tipo de muestreo por el que los sujetos se seleccionan a partir de una lista continua, eligiendo cada n número de sujetos.
- Multimétodo:** en investigación cualitativa, el uso de múltiples estrategias para corroborar los datos obtenidos a partir de cualquier estrategia simple y/o maneras de confirmar los datos dentro de una estrategia simple de recogida de datos.
- Multivariado:** familia de estadísticas que se usan cuando hay más de una variable independiente y/o más de una variable dependiente.
- Nivel de significado:** también, nivel alfa. Ver nivel de significatividad.
- Nivel de significatividad alfa:** nivel predeterminado de probabilidad para rechazar hipótesis nulas.
- Nivel de significación:** también, nivel de probabilidad, nivel de confianza. Un valor seleccionado para indicar la elección que es errónea para rechazar la hipótesis nula.
- No paramétrico:** tipos de procedimientos estadísticos que se usan cuando se violan las suposiciones necesarias para usar procedimientos paramétricos.
- No reactiva:** ver medidas discretas.
- Nominal:** tipo de escala de medida en la que los objetos o personas se nombran, clasifican o numeran.
- Notas de campo:** datos obtenidos por observación del participante en el entorno mientras el investigador esta en el campo.
- Objetividad:** procedimientos de recogida y análisis de datos a partir de los que sólo puede hacerse una interpretación o darle un significado.
- Objetivos de comportamiento:** también, objetivos de ejecución, objetivos medidos. Objetivos de una práctica definida en términos de ejecuciones terminales observables que se pueden medir.

- Observación continuada:** técnica de recogida de datos mediante la observación por la que el investigador registra todos los comportamientos importantes.
- Observación participante:** recogida de datos interactiva efectuada por los investigadores que experimentan las actividades diarias de los participantes a través de un periodo de tiempo y graban notas de campo descriptivas.
- Observaciones de campo:** técnica fundamental para cualquier investigación cualitativa en forma de detalladas notas de campo descriptivas de sucesos, personas, acciones y objetos en los entornos.
- Observaciones estructuradas:** observaciones de comportamiento sistemáticas de acuerdo a categorías específicas.
- Observador experimentado:** observador que continúa siendo completamente indiferente al grupo o al proceso de interés.
- Ordinal:** tipo de escala de medida en la que los objetos o personas se ordenan de menor a mayor.
- Paramétricos:** tipos de procedimientos estadísticos que suponen normales en las distribuciones de población, homogeneidad de varianza, e intervalo o índice de datos.
- Periodo de observación:** tipo de procedimiento de grabación del observador en el se registra el comportamiento que ocurre durante un intervalo de tiempo determinado.
- Permanencia en el campo:** (en etnografía) estar presente en el campo o en el escenario durante un tiempo largo para la recogida de datos.
- Pesos beta:** coeficiente de regresión estandarizado.
- Población:** grupo de individuos o eventos a partir de los que se confecciona la muestra.
- Polígonos de frecuencia:** representación gráfica de una distribución de frecuencia formada por la conexión en una línea de las mayores frecuencias de cada puntuación.
- Por dos mitades:** tipo de fiabilidad de consistencia interna en la que se correlacionan mitades iguales de un test.
- Posicionamiento:** investigadores cualitativos pueden mostrar sus propias posiciones o puntos de vista mediante la descripción de su propia cultura social, histórica, racial y sexual en el estudio.
- Pregunta principal:** en entrevistas cualitativas, una pregunta en la que la redacción propicia ciertas respuestas.

- Preguntas con dos objetivos:** preguntas simples que contienen dos o más ideas para las que el sujeto debe dar una respuesta.
- Preguntas estructuradas:** también, preguntas de respuesta limitada. Un tipo de pregunta de entrevista que va seguida por un conjunto de respuestas predeterminadas.
- Preguntas no estructuradas:** tipo de pregunta de entrevista que es amplia y permite respuestas abiertas.
- Preguntas semiestructuradas:** tipo de entrevista que permite respuestas individuales y de respuesta abierta a preguntas que son medianamente específicas.
- Pretest:** evaluación el resultado antes de que se administre el tratamiento.
- Probabilidad:** modo científico de definir el grado de confianza en la predicción de algo.
- Problema de investigación:** planteamiento formal de la cuestión o hipótesis que implica investigación empírica.
- Problemas previstos:** problemas de investigación anticipados que se reformularán durante la recogida de datos.
- Prueba chi-cuadrado para muestras independientes:** ver tabla de contingencia.
- Prueba de campo:** estudio preliminar de un programa o materiales, o una prueba en un entorno con probabilidad de ser utilizado.
- Prueba *t* para muestras dependientes:** también, prueba *t* de muestras correlacionadas. Un procedimiento estadístico inferencial para determinar el nivel de probabilidad de rechazar las hipótesis nulas con dos muestras de sujetos que coinciden o se relacionan.
- Prueba *t* para muestras independientes:** procedimiento inferencial estadístico para determinar el nivel de probabilidad o rechazo de las hipótesis nulas con dos muestras de sujetos que no tienen relación con los demás.
- Prueba *t*:** procedimiento estadístico inferencial para la determinación del nivel de probabilidad del rechazo de la hipótesis nula que es igual en dos medias.
- Pruebas alternativas:** procedimientos para medir la ejecución a través de preguntas de respuesta cerrada a diferencia de los tests de lápiz y papel tradicionales.
- Pruebas de lápiz y papel:** instrumentos que contienen un conjunto de preguntas estructuradas que miden los niveles y el conocimiento cognitivo.
- Pruebas típicas:** pruebas que se administran y puntúan de acuerdo a instrucciones prescritas y altamente estructuradas.
- Puntuaciones T:** tipo de puntuación típica con una media de cincuenta y una derivación diez.

Puntuaciones típicas: número convertidos a partir de distribuciones puras con medias constantes y derivaciones típicas.

Puntuaciones z: tipo de puntuación estándar con una media de cero y una derivación típica de uno.

Semiexperimental: investigación que no tiene asignación al azar de sujetos, pero investiga relaciones causa-efecto mediante la manipulación de la variable independiente.

Rango de percentil: punto en una distribución en el que o por debajo de él se encuentra un porcentaje de puntuaciones determinado.

Rango: medida de variabilidad; la diferencia entre las mayores y menores puntuaciones en una distribución.

Razonamiento deductivo: razonamiento que asume que si las premisas son correctas, automáticamente la conclusión es correcta.

Razonamiento inductivo: razonamiento por el que la observación de casos particulares se generaliza a la clase completa.

Recuento de frecuencias: tipo de procedimiento de grabación de observador en el que se graba la frecuencia del comportamiento.

Referencia a criterio: instrumentos cuyas puntuaciones se interpretan por referencia a unos criterios o estándares específicos además de por la ejecución de otros.

Referido a normas: interpretación de los resultados de un test en la que una puntuación o grupo de puntuaciones se compara con la ejecución típica de un grupo determinado (normal).

Reflexión: autoescrutinio riguroso del investigador, desde el principio hasta el fin del proceso de investigación cualitativa.

Registros cualitativos de campo: datos grabados como notas de campo de observación del participante, registros de entrevistas en profundidad, o notas del investigador de documentos históricos.

Registro de duración: tipo de registro de observación que consigna la duración del comportamiento.

Registro de frecuencia: tipo de registro de observación que consigna la frecuencia del comportamiento.

Registros inmediato: registros efectuados inmediatamente después de abandonar el campo; contienen resumen de observaciones, tratamiento de la calidad de los datos, sugerencias de pasos siguientes, y notas de autoseguimiento.

Registro de intervalo: tipo de registro de observación que consigna el comportamiento que sucede durante intervalos de tiempo.

- Regresión estadística:** tendencia para que puntuaciones extremas lleguen a ser más próximas a la puntuación media en una segunda comprobación.
- Relaciones espúreas:** conexión que sobrerrepresenta o infrarepresenta una relación real.
- Relación negativa:** relación en la que el incremento de una variable se corresponde con la disminución de otra.
- Relación positiva:** relación en la que el incremento de las puntuaciones de una variable se corresponde con el incremento de las puntuaciones de otra.
- Replica:** estudio que duplica un estudio previo pero que usa diferentes ambientes o técnicas.
- Repeticiones del tratamiento:** amenaza a la validez interna en la que el número de replicaciones del tratamiento no es igual al número de sujetos.
- Respuesta abierta:** tipo de ítem del cuestionario en el que los sujetos escriben la respuesta a una pregunta.
- Respuesta cerrada:** tipo de ítem del cuestionario en el que el sujeto elige entre opciones predeterminadas.
- Respuesta múltiple:** tipo de elemento de cuestionario en el que los sujetos marcan las repuestas apropiadas que se les proporcionan.
- Restricción del rango:** conjunto de puntuaciones que representa sólo una parte de la distribución total.
- Segmento:** fragmento del conjunto de datos que es comprensible por sí misma y contiene una idea, episodio o sección de información relevante del estudio.
- Selección del escenario:** especificación de los criterios de escenario implicados en los problemas previstos y para la obtención de un escenario factible y adecuado.
- Selección:** amenaza a la validez interna en que las diferencias entre grupos o sujetos afectan a los resultados.
- Significación del problema:** base fundamental de un problema o la importancia de un estudio en relación con el desarrollo de la teoría, conocimiento, y/o práctica educativa.
- Síntesis de la investigación:** procedimiento que evalúa sistemáticamente o resume estadísticamente o en forma de narración estudios comparables.
- Sondeo de las entrevistas:** breves cuestiones o frases que obtiene detalles para su elaboración, explicaciones adicionales y aclaración de las respuestas.

Subjetividad disciplinada: autoseguimiento riguroso del investigador, esto es, replanteamiento continuo y reevaluación de todas las fases del proceso de investigación.

Sujetos: persona o personas a partir de las cuales se recogen los datos en un estudio.

Tabla de contingencias: también, prueba chi-cuadrado para muestras independientes. Una prueba chi-cuadrado con dos o más variables independientes.

Tamaño del efecto: índice estadístico de la diferencia práctica o estadística entre grupos.

Tanteo: ver validación de la entrevista.

Técnicas cualitativas: estrategias de recogida de datos que adquieren los datos en forma de palabras en lugar de números.

Técnicas cuantitativas: técnicas de recogida de datos que usan números para describir o medir los resultados.

Técnicas suplementarias: técnicas cualitativas que ayudan a interpretar, elaborar o corroborar los datos obtenidos a partir de la observación del participante, entrevistas en profundidad, documentos, y artefactos.

Tema: Nombre descriptivo para los asuntos de un sujeto de un segmento.

Teoría: predicción y explicación de los fenómenos naturales.

Teorías argumentadas o fundamentadas: procedimientos cualitativos para desarrollar conceptos detallados o proposiciones condicionales para una teoría sustantiva.

Test de aptitud: test que se usa para predecir el comportamiento.

Test de habilidad o de rendimiento: tests que miden el conocimiento, los niveles o el comportamiento.

Tests estandarizados: tests que son administrados y puntuados según una fuerte estructura e instrucciones prescritas.

Tests de rendimiento: tests que miden el conocimiento, la destreza o el comportamiento.

Testimonios orales: registros o transcripciones de entrevistas de testigos o participantes del evento pasado que se está estudiando.

Tipicidad: grado en que un fenómeno puede compararse o contrastarse con otros fenómenos con respecto a dimensiones relevantes.

Tratamiento de datos: uso de un sistema para recuperar los conjuntos de datos y ensamblar los datos codificados en un sitio.

- Triangulación:** validación cruzada cualitativa entre múltiples fuentes de datos, estrategias de recogida de datos, periodos de tiempo, y esquemas teóricos.
- Validez del constructo:** tipo de validez externa que se refiere al grado en que el estudio representa el constructo fundamental.
- Validez del contenido:** tipo de validez para validar en qué el contenido de un test se juzga representativo de un gran dominio de contenido.
- Validez del test:** grado en el que son apropiadas las inferencias realizadas sobre la base de las puntuaciones del test.
- Validez de diseños cualitativos:** grado en el que las interpretaciones y conceptos tienen los mismos significados entre los participantes y el investigador.
- Validez externa:** se refiere al grado con el que los resultados de un estudio pueden generalizarse a otros sujetos, condiciones o situaciones.
- Validez interna:** grado en que se controlan las variables extrañas.
- Validez:** grado en el que las explicaciones científicas de los fenómenos coinciden con el mundo real.
- Variabilidad:** ver medidas de variabilidad.
- Variable categórica:** variable utilizada para dividir los sujetos, objetos o entidades en dos o más grupos.
- Variable continua:** también, variable medida. Una variable en la que la propiedad o el atributo de un objeto, sujeto, o entidad se mide de forma numérica y un rango puede asumir un número infinito de valores.
- Variable criterio:** en un estudio predictivo, la variable que se predice.
- Variable dependiente:** variable medida que es consecuencia de o depende de las variables antecesoras.
- Variable experimental:** variable en el diseño experimental o semiexperimental que el investigador manipula o cambia para ver el efecto sobre (su relación con) la variable dependiente.
- Variable independiente:** variable que antecede o precede a la variable dependiente; en diseño experimental también se denomina variable experimental o manipulada.
- Variable manipulada:** en los experimentos la variable independiente que se determina por el investigador.
- Variable medida:** ver variable continua.
- Variable predictora:** variable antecedente en un estudio de predicción.

Variable: acontecimiento, categoría, comportamiento, o un atributo que expresa un constructo y posee diferentes valores, dependiendo de cómo se utilice en un estudio.

Variables asignadas: ver variables atribuidas.

Variables atribuidas: también, variables asignadas. Variables independientes que no pueden manipularse.

Varianza residual del constructo: grado en el que un acuerdo mide las facetas no relacionadas con el propósito del acuerdo.

Varianza: generalmente, el grado de difusión o dispersión de las puntuaciones, matemáticamente, el cuadrado de la derivación típica.

Verificación: resultados de la investigación que pueden confirmarse o modificarse en posteriores investigaciones.

Cálculos para estadísticos descriptivos e inferenciales seleccionados

En este apéndice presentamos una guía detallada para el cálculo de varios procedimientos estadísticos simples¹. Nuestra intención no es deducir fórmulas, sino mostrar cómo se calculan los estadísticos. Creemos que ser capaces de aplicar estas fórmulas ayuda en gran medida a la comprensión del significado de las estadísticas.

MEDIDAS DE TENDENCIA CENTRAL

Las medidas de tendencia central son estadísticos descriptivos que miden el valor o la localización central de conjuntos de puntuaciones. Se usan ampliamente para resumir y simplificar grandes cantidades de datos.

La media

La media es el promedio aritmético de un conjunto de puntuaciones. Se obtiene sumando todas las puntuaciones de una distribución y dividiendo la suma entre el número de puntuaciones.

La fórmula para calcular la media es:

$$\bar{X} = \frac{\Sigma X}{n}$$

donde

\bar{X} es la puntuación media

ΣX es la suma de las X_s (por ejemplo, $X_1 + X_2 + X_3 \dots X_n$)

n es el número total de puntuaciones

Ejemplo: Cálculo de la media. Si hemos obtenido una muestra de ocho puntuaciones -17, 14, 14, 13, 10, 8, 7, 7- la media de este conjunto de puntuaciones se calcula

$$\begin{aligned} \Sigma X &= 17 + 14 + 14 + \dots + 7 = 90 \\ n &= 8 \end{aligned}$$

¹ Las tablas estadísticas se encuentran al final del apéndice.

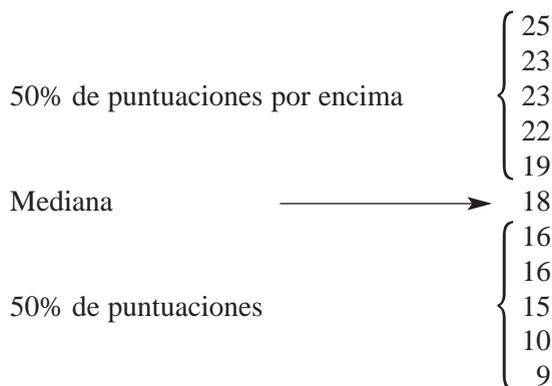
Por lo tanto,

$$\bar{X} = \frac{90}{8} = 11,25$$

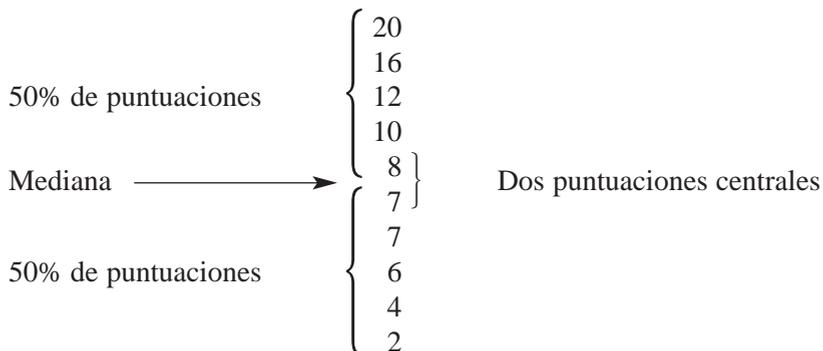
La mediana

La mediana es la puntuación de una distribución bajo la cual están la mitad de las puntuaciones. En otras palabras, la mitad de las puntuaciones está por encima de la mediana y la mitad está por debajo. La mediana está en el percentil 50.

Para calcular la mediana, las puntuaciones se ordenan de mayor a menor; después simplemente se cuenta, desde un extremo, la mitad de las puntuaciones. En distribuciones con un número impar de puntuaciones la mediana es la puntuación que está en el punto medio como se ilustra a continuación:



Si la distribución tiene un número par de puntuaciones, la mediana es la media de las dos puntuaciones que están en el punto medio. En este caso la mediana es un nuevo punto o puntuación en la distribución como se muestra a continuación:



De esa manera, la mediana en este ejemplo es:

$$7 + \frac{8}{2} = \frac{15}{2} = 7,5$$

La mediana es relativamente fácil de encontrar en estudios con un número pequeño de sujetos. Conforme se incrementa el número de puntuaciones, el cálculo se hace mediante una fórmula o por la colocación de las puntuaciones en intervalos de puntuaciones y usando el intervalo para hacer el cálculo. El ordenador es capaz de hacer estos cálculos más complicados de forma fácil, rápida y fiable.

La moda

La moda es simplemente la puntuación más frecuente en una distribución y se halla contando el número de veces que aparece cada puntuación. En la distribución que mostramos a continuación la moda es 22.

23
22
22
22
20
18
18
17
16

MEDIDAS DE VARIABILIDAD

Las medidas de variabilidad se utilizan para mostrar las diferencias entre las puntuaciones de una distribución. Usamos el término *variabilidad* o *dispersión* debido a que los estadísticos proporcionan una indicación de lo diferentes o dispersas que son unas puntuaciones de otras. Presentaremos tres medidas de variabilidad: rango, varianza y desviación típica.

El rango

El rango es la medida de variabilidad más simple, pero también la menos útil. El rango se define como la distancia entre la puntuación más alta y la más baja y se calcula simplemente restando la puntuación más baja de la puntuación más alta:

$$\text{Rango} = X_A - X_B$$

donde

X_A = la puntuación más alta

X_B = la puntuación más baja

Por lo tanto, para los siguientes resultados, el rango es: $26 - 6 = 20$.

6 8 10 11 15 20 26

El rango es una medida de variabilidad imperfecta e inestable. Debido a que el rango puede estar sesgado, raramente se usa como la única medida de variabilidad.

La varianza

La varianza (s^2 o σ^2) es una medida de dispersión que indica el grado en el que se agrupan las puntuaciones alrededor de la media. La varianza proporciona al investigador un número para indicar, en cierto sentido, la dispersión media de las puntuaciones desde la media. Desde el punto de vista computacional, la varianza es la suma de las desviaciones al cuadrado dividido entre el número total de puntuaciones.

$$s^2 = \frac{\sum(X - \bar{X})^2}{N}$$

donde

s^2 es la varianza

$\sum(X - \bar{X})^2$ es la suma de las desviaciones al cuadrado

$(X - \bar{X})$ es la desviación

N es el número total de puntuaciones

Para cualquier distribución de puntuaciones, la varianza puede determinarse por los siguientes cinco pasos:

1. Calcular la media: $(\sum X/N)$.
2. Calcular la desviación: $(X - \bar{X})$.
3. Elevar al cuadrado cada desviación: $(X - \bar{X})^2$.
4. Sumar todas las puntuaciones de desviación al cuadrado: $\sum(X - \bar{X})^2$.
5. Dividir la suma entre N : $\sum(X - \bar{X})^2/N$.

Estos cinco pasos se ilustran en forma numérica como sigue:

(1) Puntuaciones directas	(2) $(X - \bar{X})$	(3) $(X - \bar{X})^2$	(4)	(5)
20	7	49	$\Sigma(X - \bar{X})^2 = 120$	$\frac{\Sigma(X - \bar{X})^2}{N} = 12$
15	2	4		
15	2	4		
14	1	1		
14	1	1		
14	1	1		
12	-1	1		
10	-3	9		
8	-5	25		
8	-5	25		
$\Sigma X = 130$				
$N = 10$				
$\bar{X} = 13$				

Sustituyendo directamente en la fórmula:

$$s^2 = \frac{120}{10} = 12$$

Otra fórmula, más sencilla desde el punto de vista computacional, que puede utilizarse para calcular la varianza, es:

$$s^2 = \frac{\Sigma X^2 - N\bar{X}^2}{N}$$

Debido a que la varianza se expresa como el cuadrado de las puntuaciones directas, no de las unidades originales, generalmente, no se presenta en la investigación. Para volver a unidades que sean compatibles con la distribución de puntuaciones directas, necesitamos obtener la raíz cuadrada de la varianza. Tomando la raíz cuadrada de la varianza conseguimos la desviación típica.

Desviación típica

La desviación típica (s , σ o DT) es la raíz cuadrada de la varianza. Es una medida de dispersión que emplea las puntuaciones de desviación expresadas en unidades estandarizadas a la media; de ahí el nombre de desviación típica. La desviación típica es igual a la raíz cuadrada de la suma de las puntuaciones de desviación a la media al cuadrado divididas entre el número total de puntuaciones. La fórmula es:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$

donde

s es la desviación típica

$\sqrt{\quad}$ es la raíz cuadrada

$\sum(X - \bar{X})^2$ es la suma de la desviación al cuadrado

$(X - \bar{X})$ es la desviación

N es el número total de puntuaciones

Para calcular la desviación típica, sólo hay que añadir un paso a la fórmula de la varianza: obtener la raíz cuadrada. Por ejemplo, en nuestro ejemplo para la varianza, la desviación típica sería:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{120}{10}} = \sqrt{12} = 3,46$$

La desviación típica se presenta habitualmente en la investigación y, junto con la media, es el estadístico más importante en investigación. Informa del número de puntuaciones (es decir, el porcentaje de puntuaciones) que están dentro de las unidades dadas de la desviación típica alrededor de la media. Esta propiedad de la desviación típica se explica en la sección denominada «Distribución normal».

PUNTUACIONES TÍPICAS

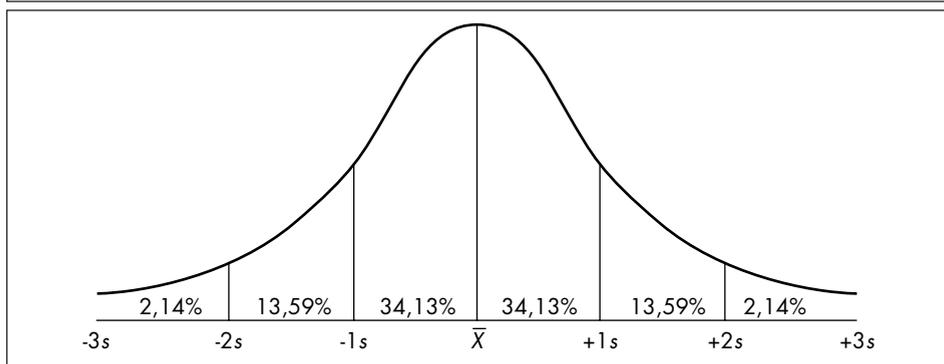
Las puntuaciones típicas son números que se han transformado a partir de las puntuaciones directas para proporcionar información coherente acerca de la localización de una puntuación dentro de una distribución total. Hay números que están relacionados con la distribución normal.

Distribución normal

La distribución normal es un conjunto de puntuaciones que, cuando se expresan gráficamente en una distribución de frecuencias, tiene como resultado una curva simétrica con forma de campana y con unas propiedades matemáticas precisas. Las propiedades matemáticas proporcionan la base para la realización de interpretaciones estandarizadas. Estas propiedades incluyen poseer una moda, una media y una mediana que sean iguales; tener una media que divida la curva en dos mitades idénticas y contar con medidas de desviación típica que se sitúen en lugares previsibles dentro de la curva normal, con el mismo porcentaje de puntuaciones entre la media y los puntos equidistantes a ésta. Esta tercera característica es muy importante. Sabemos, por ejemplo, que en $+1s$, siempre estaremos aproximadamente en el percentil 84 de la distribución (el percentil es el porcentaje de puntuaciones en o por

debajo de la puntuación designada). Esto se debe a que la mediana está en el percentil 50 y +1s contiene un 34% de puntuaciones adicional (50 + 34 = 84). De forma similar, el porcentaje de puntuaciones entre +1s y +2s es aproximadamente el 14%, lo que significa que +2s está en el percentil 98. Estas características se ilustran en la figura C.1, el gráfico de la distribución normal típica.

FIGURA C.1: GRAPH OF THE STANDARD NORMAL DISTRIBUTION OR NORMAL CURVE



Lo mejor de esta propiedad es que, para cualquier distribución de puntuaciones directas con unidades específicas, como por ejemplo 1 o 2 como s , la interpretación es siempre la misma. Por lo tanto, si una distribución tiene una media de 10 y una desviación típica de 3, y una segunda distribución una media de 50 y una desviación típica de 7, una puntuación de 4 en el primer caso está aproximadamente en el mismo percentil (el segundo) que una puntuación de 36 en la segunda distribución.

Puntuaciones z

La puntuación típica más básica se denomina puntuación z y se expresa como una desviación de la media en unidades de desviación típica. Una puntuación z de 1 es, por lo tanto, un desviación típica, -1 es una menor desviación típica, $+2$ en dos desviaciones típicas, y así sucesivamente.

Después de calcular la media y la desviación típica para un conjunto de puntuaciones, es fácil convertir cada puntuación directa en una puntuación z , con la que indicaremos exactamente donde se sitúa exactamente cada puntuación en la distribución normal.

La fórmula para calcular una puntuación z es:

$$z = \frac{X - \bar{X}}{s}$$

donde

z es el valor de la puntuación z

X es cualquier puntuación

\bar{X} es la media aritmética de una distribución de puntuaciones

s es la desviación típica de la misma distribución

Tomando las puntuaciones utilizadas para ejemplificar la varianza y la desviación típica, las puntuaciones típicas se calcularían como sigue:

$$\text{Para la puntuación directa de 20: } z = \frac{20 - 13}{3,46} = 2,02$$

$$\text{Para la puntuación directa de 14: } z = \frac{14 - 13}{3,46} = 0,29$$

$$\text{Para la puntuación directa de 10: } z = \frac{1 - 13}{3,46} = 0,87$$

Una vez se han calculado las puntuaciones z , es fácil acudir a las tablas de conversión para encontrar el percentil correspondiente a cada puntuación z .

Puntuaciones t

Una limitación del uso de puntuaciones z es la necesidad de tener cuidado con los signos negativos y con el punto decimal. Para evitar estos problemas, se usan otras puntuaciones típicas para convertir algebraicamente las puntuaciones z a unidades diferentes. La fórmula general para convertir las puntuaciones z es

$$A = \bar{X}_A + s_A(z)$$

donde

A es la nueva puntuación típica equivalente a z

\bar{X}_A es la media para la nueva escala de puntuaciones típicas

s_A es la desviación típica de la nueva escala

z es la puntuación z para cualquier observación

Para las puntuaciones t , $\bar{X}_A = 50$ y $s_A = 10$, la fórmula para la conversión de puntuaciones z en puntuaciones t es

$$t = 50 + 10(z)$$

Por ejemplo, las puntuaciones t para nuestro supuesto anterior serían:

$$\text{Para la puntuación directa de 20: } t = 50 + 10(2.02) = 70,2$$

$$\text{Para la puntuación directa de 14: } t = 50 + 10(0.29) = 52,9$$

$$\text{Para la puntuación directa de 10: } t = 50 + 10(-0.87) = 41,3$$

Otras puntuaciones típicas

Otras puntuaciones típicas frecuentes son las siguientes:

1. Equivalente de curva normal (ECN) tiene una media de 50 y s de 21,06. De ese modo $ECN = 50 + 21,06$ (puntuación z).
2. Puntuación CI tiene una media de 100 y s de 15 o 16. De ese modo $CI = 100 + 15$ (puntuación z).
3. *College Entrance Examination Boards* (CEEB, como por ejemplo SAT) usa una media de 500 y una s de 100. De ese modo, $CEEB = 500 + 100$ (puntuación z).
4. ACT (*American College Testing Program*) usa una media de 20 y una s de 5. De ese modo, $ACT = 20 + 5$ (puntuación z).
5. Estaninos. Los estaninos también se presentan normalmente. El estanino está tipificado en una media de 5 y una s de 2, pero, al contrario que otras puntuaciones típicas, los números hacen referencia a intervalos en lugar de a puntos determinados de la distribución normal. El estanino 5 se localiza en el centro de la distribución e incluye el 20% medio de las puntuaciones; los estaninos 4 y 6 incluyen el 17% de las puntuaciones; los 3 y 7, el 12%; los 2 y 8, el 7%; y los 1 y 9, el 4%. Los estaninos se ilustran en la figura 6.11, la curva normal.

MEDIDAS DE RELACIÓN

Las medidas de relación se emplean para indicar el grado en el que se relacionan o covarían dos conjuntos de puntuaciones. Intuitivamente buscamos relaciones mediante planteamientos como: «si puntuaciones altas de una variable X tienden a asociarse con puntuaciones altas de la variable Y , entonces las variables están relacionadas», o «si puntuaciones altas de una variable X tienden a asociarse con puntuaciones bajas de la variable Y , entonces las variables están relacionadas». Como indicamos en el capítulo 4, la relación puede ser tanto positiva como negativa, así como fuerte o débil.

Utilizamos coeficientes de correlación como un resumen estadístico de la naturaleza de la relación entre dos variables. Nos proporcionan una estimación del grado cuantitativo de relación. Los números están casi siempre entre $-1,00$ y $+1,00$. Mostraremos cómo calcular los dos coeficientes de correlación más comunes, producto-momento de Pearson y los rangos de Spearman (ρ).

Producto-momento de Pearson (r de Pearson)

El coeficiente de correlación de producto-momento de Pearson es la medida de relación más ampliamente utilizada. La r de Pearson se calcula para mostrar la relación lineal entre dos variables. Para calcular la r de Pearson, se necesitan dos medidas de cada sujeto. Por ejemplo, suponga que tenemos un grupo de diez

sujetos y para cada sujeto tenemos medidas de autoestima y de rendimiento. Podemos calcular la r de Pearson entre el autoconcepto y el rendimiento para estos diez sujetos usando la siguiente fórmula:

$$r \text{ de Pearson} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \cdot \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$$

donde

ΣXY es la suma de los productos cruzados XY

ΣX es la suma de las puntuaciones X

ΣY es la suma de las puntuaciones Y

ΣX^2 es la suma de las puntuaciones X al cuadrado

ΣY^2 es la suma de las puntuaciones Y al cuadrado

N es el número de pares de puntuaciones

Esta fórmula puede parecer compleja pero en realidad es bastante fácil de calcular. Estas puntuaciones pueden listarse en una tabla, como se muestra a continuación; para utilizarlas, simplemente se hallan los valores para cada sumatorio en la fórmula, se sustituye donde sea apropiado y se realizan las operaciones indicadas.

Sujeto	Puntuación autoestima X	X^2	Puntuación rendimiento Y	Y^2	$X \cdot Y$
1	25	625	85	7.225	2.125
2	20	400	90	8.100	1.800
3	21	441	80	6.400	1.680
4	18	324	70	4.900	1.260
5	15	225	75	5.625	1.125
6	17	289	80	6.400	1.360
7	14	196	75	5.625	1.050
8	15	225	70	4.900	1.050
9	12	144	75	5.625	900
10	13	169	60	3.600	780
	$\Sigma X = 170$	$\Sigma X^2 = 3.038$	$\Sigma Y = 760$	$\Sigma Y^2 = 58.400$	$\Sigma X \cdot Y = 13.130$
	$(\Sigma X)^2 = 28.900$		$(\Sigma Y)^2 = 577.600$		

Paso 1: Emparejar cada conjunto de puntuaciones; un conjunto es X , el otro es Y .

Paso 2: Calcular ΣX y ΣY .

Paso 3: Calcular X^2 e Y^2 .

Paso 4: Calcular ΣX^2 e ΣY^2 .

Paso 5: Calcular $(\Sigma X)^2$ e $(\Sigma Y)^2$

Paso 6: Calcular $X \times Y$.

Paso 7: Calcular $\Sigma X \times Y$.

Paso 8: Sustituir los valores calculados en la fórmula.

$$\begin{aligned}
 r \text{ de Pearson} &= \frac{10(13.130) - (170)(760)}{\sqrt{10(3.038) - 28.900} \cdot \sqrt{10(58.400) - 577.600}} \\
 &= \frac{13.130 - 129.200}{\sqrt{3.038 - 28.900} \cdot \sqrt{58.400 - 577.600}} \\
 &= \frac{2.100}{\sqrt{1.480} \cdot \sqrt{6.400}} \\
 &= \frac{2.100}{38,47 \cdot 80} \\
 &= \frac{2.100}{3.078} \\
 &= 0,68
 \end{aligned}$$

El valor de 0,68 muestra una relación positiva moderada entre la autoestima y el rendimiento para este conjunto de puntuaciones. El nivel de significación de los coeficientes de correlación se indica en la tabla C.2 al final de este apéndice.

Rangos de Spearman (rangos r o rho de Spearman)

La rho de Spearman se utiliza cuando se dispone de rangos en dos variables para todos los sujetos. Los rangos son listados de puntuaciones de mayor a menor. La correlación rho de Spearman muestra el grado con el que los sujetos mantienen la misma posición relativa en dos medidas. En otras palabras, la rho de Spearman indica cuanto acuerdo existe entre los rangos de cada variable.

El cálculo de los rangos de Spearman es más fácil que el cálculo de la r de Pearson. Los pasos son:

Paso 1: Ordenar las Xs e Ys.

Paso 2: Emparejar la ordenación de Xs e Ys.

Paso 3: Calcular la diferencia en los rangos para cada par.

Paso 4: Elevar al cuadrado cada diferencia.

Paso 5: Sumar las diferencias al cuadrado.

Paso 6: Sustituir los valores calculados en la fórmula.

La fórmula es:

$$\text{rho de Spearman} = 1 - \frac{6\sum D^2}{n(n^2 - 1)}$$

Para los datos que se usaron en el cálculo de la r de Pearson, la rho de Spearman se hallaría del siguiente modo:

Sujeto	Rango autoestima X	Rango rendimiento Y	Diferencia D	D^2
1	1	2	-1	1
2	3	1	2	4
3	2	3,5	-1,5	2,25
4	4	5,5	-1,5	2,25
5	6,5	8	-1,5	2,25
6	5	3,5	1,5	2,25
7	8	8	0	0
8	6,5	5,5	1	1
9	10	8	2	4
10	9	10	-1	1
				$\sum D^2 = 20$

Nota: Cuando existen empates en la ordenación, todas las puntuaciones empatadas reciben el promedio de los rangos implicados.

$$\begin{aligned} \text{rangos } r &= \frac{1 - 6(20)}{10(100 - 1)} \\ &= 1 - \frac{120}{990} \\ &= 1 - 0,12 \\ &= 0,88 \end{aligned}$$

En la mayoría de los conjuntos de datos con más de cincuenta sujetos, la r de Pearson y el rango de Spearman proporcionarán correlaciones casi idénticas. En el ejemplo presentado aquí, la rho de Spearman es mayor debido al bajo tamaño de n y a que los tipos de empates en los rangos dan como resultado bajas puntuaciones de diferencia.

CHI-CUADRADO

Chi-cuadrado (χ^2) es un procedimiento estadístico que se usa como un estadístico inferencial con datos nominales como recuentos de frecuencias y con datos ordinales como porcentajes y proporciones. En el caso más simple, los datos se organizan en dos categorías, como por ejemplo sí y no, alto y bajo, a favor y en contra. Si un investigador, por ejemplo, está interesado en la opinión de los profesores universitarios sobre el contrato indefinido y pregunta: «¿Debería suprimirse la contratación indefinida?», entonces todas las respuestas podrían separarse en categorías como *sí* o *no*. La frecuencia total de cada categoría (frecuencias observadas) se compara con la frecuencia esperada, lo que en la mayoría de los casos es casual. Esto significa que, con dos categorías, la mitad de las respuestas deberían ser *sí* y la otra mitad *no*. Si asumimos los siguientes resultados:

¿Debería suprimirse la contratación indefinida?

	Sí	No
Observada	40	60
Esperada	50	50

Estos valores se usan en la siguiente fórmula para calcular el estadístico chi-cuadrado:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

donde

χ^2 es el estadístico chi-cuadrado

Σ es el sumatorio

f_o es la frecuencia observada

f_e es la frecuencia esperada

Insertando los valores de la tabla, el resultado es:

$$\begin{aligned} \chi^2 &= \frac{(40 - 50)^2}{50} + \frac{(60 - 50)^2}{50} \\ &= \frac{100}{50} + \frac{100}{50} \\ &= 2 + 2 \\ &= 4,0 \end{aligned}$$

El valor obtenido, en este caso 4, se utiliza con los grados de libertad en el problema ($gl = c - 1$, donde c equivale al número de categorías; en nuestro ejemplo $gl = 2 - 1$, o 1) para hallar el valor de chi-cuadrado en los valores críticos de la tabla chi-cuadrado (tabla C.3 al final del apéndice) para determinar el nivel de significación de los resultados. En referencia a la tabla y localizando 4.00 dentro de la tabla con 1 gl, el resultado es significativo con un valor p de 0.05. Consecuentemente, sería apropiado decir que hay una diferencia significativa en el número de profesores que responden *sí* comparado con el número que responden *no*.

Suponga que el investigador quisiera avanzar un paso más en este problema y conocer si los profesores y directores difieren en sus respuestas acerca de la supresión de la contratación indefinida. El investigador debería tener lo que se llama una tabla de contingencia, que es una tabulación cruzada de las frecuencias para las combinaciones de categorías de las dos variables. A continuación se muestra una tabla de contingencia hipotética para profesores y directores.

¿Debería suprimirse la contratación indefinida?

	Profesores	Directores	Totales
Sí	40($p = 0.40$)	40($p = 0.80$)	80($p_e = 0.53$)
No	60($p = 0.60$) $n = 100$	10($p = 0.20$) $n = 50$	70($p_e = 0.47$) $n = 150$

Observe que en la tabla, la proporción de respuestas dentro de cada categoría (*sí* y *no*) se muestra tanto para profesores como para directores y que en la última columna se exponen las proporciones totales. Estas proporciones se usan en la siguiente fórmula:

$$\chi^2 = \frac{\sum n(P - P_e)^2}{P_e}$$

donde

χ^2 es el estadístico chi-cuadrado

\sum es la suma de todas las celdas en el problema (en nuestro ejemplo hay cuatro celdas)

n es el número total de observaciones en cada columna

P es la proporción de frecuencias observadas en cada celda

P_e es la proporción esperada para cada fila

Por lo tanto, para nuestro ejemplo el resultado sería:

$$\chi^2 = 100 \frac{(0.4 - 0.53)^2}{0.53} + 100 \frac{(0.06 - 0.47)^2}{0.47}$$

$$\begin{aligned}
 &= 50 \frac{(0.80 - 0.53)^2}{0.53} + 100 \frac{(0.20 - 0.47)^2}{0.47} \\
 &= 100 \frac{0.20}{0.53} + 100 \frac{0.20}{0.47} + 50 \frac{0.07}{0.53} + 50 \frac{0.07}{0.47} \\
 &= 3.77 + 4.26 + 6.60 + 7.45 \\
 &= 22.08
 \end{aligned}$$

En las tablas de contingencia los grados de libertad equivalen a $(r - 1)(c - 1)$, donde r es el número de filas y c el número de columnas. En nuestro ejemplo los $gl = (2 - 1)(2 - 1) = 1$. Localizando 22.08 con 1 grado de libertad en los valores críticos de la tabla chi-cuadrado (tabla C.3) advertimos que el resultado es altamente significativo, $p < 0.001$. Este resultado indica que existe una asociación o relación significativa entre las dos variables (profesores y directores y sí y no).

PRUEBA TEST *t*

La prueba *t* se utiliza para indicar la probabilidad de que las medias de dos grupos sean diferentes. Presentaremos dos formas habituales de la prueba *t*, una se emplea con muestras independientes y la otra con muestras dependientes.

Prueba *t* para muestras independientes

La prueba *t* para muestras independientes, o prueba *t* para grupos independientes, se emplea para determinar si los valores de la media de una variable en un grupo de sujetos es diferente al valor de la media de la misma variable en un grupo diferente de sujetos. Es importante cumplir tres condiciones estadísticas: que las distribuciones de frecuencia de las puntuaciones del conjunto de individuos a los que se realiza la muestra de cada grupo sean normales; que las varianzas en cada conjunto de individuos a los que se realiza la muestra sean iguales y que la observación de las puntuaciones en un grupo sea independiente de la del otro grupo. Si el tamaño de la muestra es mayor que 30, el incumplimiento de la condición de la normalidad no es grave; y mientras que los tamaños de las muestras sean iguales, el incumplimiento de la suposición de la homogeneidad no es un problema. Sin embargo, resulta crucial que las observaciones de cada grupo sean independientes.

La fórmula para el cálculo del estadístico *t* es:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

donde

t es el estadístico t

\bar{X}_1 es la media de un grupo

\bar{X}_2 es la media del segundo grupo

$s_{\bar{x}_1 - \bar{x}_2}$ es el error típico de la diferencia de las medias

El error típico de la diferencia de las medias se estima a partir de las varianzas de cada distribución. Esta parte de la fórmula se calcula combinando las varianzas de cada distribución para dar lugar a s , lo que se hace mediante la siguiente fórmula:

$$s = \sqrt{\frac{\sum x_1^2 + \sum x_2^2}{gl_1 + gl_2}}$$

Entonces,

$$s_{\bar{x}_1 - \bar{x}_2} = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Tomemos los siguientes datos como un ejemplo:

Grupo x_1	Grupo x_2
$\bar{X}_1 = \bar{X}_1 18$	$\bar{X}_2 = 25$
$n_1 = 20$	$n_2 = 20$
$\sum x_1^2 = 348$	$\sum x_2^2 = 425$

A partir de este punto podemos calcular el estadístico t en los siguientes pasos

1. Calcular s :

$$\begin{aligned} s &= \sqrt{\frac{348 + 425}{19 + 19}} \\ &= \sqrt{20.34} \\ &= 4.51 \end{aligned}$$

2. Calcular $s_{\bar{x}_1 - \bar{x}_2}$

$$\begin{aligned} s_{\bar{x}_1 - \bar{x}_2} &= 4.51 \sqrt{\frac{1}{20} + \frac{1}{20}} \\ &= 4.51 \sqrt{\frac{1}{10}} \\ &= 4.51 (0.32) \\ &= 1.44 \end{aligned}$$

Sustituir en la fórmula de la prueba t :

$$\begin{aligned}
 t &= \frac{18 - 25}{1.44} \\
 &= \frac{7}{1.44} \\
 &= 4.86
 \end{aligned}$$

Una vez se ha calculado el estadístico t , se busca en los valores críticos para la tabla de la prueba t (tabla 11.1) con sus correspondientes grados de libertad (que para la prueba t para muestras independientes es $n_1 + n_2 - 2$ ó, en nuestro ejemplo, $20 + 20 - 2 = 38$) con el objeto de determinar el nivel de significación de los resultados. En este ejemplo, el estadístico t de 4.86, con 38 gl , es significativa con una $p < 0.001$.

Aquí hay otro ejemplo de cálculo con la prueba t , comenzando con los datos directos:

Grupo 1		Grupo 2	
x_1	x_1^2	x_2	x_2^2
		7	49
8	64	7	49
8	64	8	64
6	36	6	36
5	25	6	36
5	25	4	16
6	36	4	16
6	36	3	9
9	81	5	25
8	64	5	25
$\Sigma x_1 = 68$	$\Sigma x_1^2 = 480$	$\Sigma x_2 = 55$	$\Sigma x_2^2 = 325$
$n = 10$	$n = 10$	$\bar{X}_2 = 5.5$	
$\bar{X}_1 = 6.8$			

Paso 1: $s = \sqrt{\frac{480 + 325}{9 + 9}} = \sqrt{44.72} = 6.69$

Paso 2: $s_{\bar{x}_1 - \bar{x}_2} = 6.69 \sqrt{\frac{1}{10} + \frac{1}{10}} = 6.69 \sqrt{\frac{1}{5}} = (6.69)(0.45) = 2.99$

Paso 3: $t = \frac{6.8 - 5.5}{2.99} = \frac{1.3}{2.99} = 0.43$

En este caso, el estadístico t de 0.43, con 18 gl no es estadísticamente significativo. Por lo tanto, incluso aunque las medias de los grupos son diferentes, existe una posibilidad de que lo sean sólo por casualidad.

Prueba t para muestras dependientes

Cuando dos grupos que han sido emparejados son comparados, como en el diseño pretest-postest, la fórmula de la prueba t debe tener en cuenta la interrelación entre los grupos; es decir, que los grupos no son independientes; más aún, están relacionados. La fórmula para este tipo de prueba t es más fácil de calcular que para la prueba t de muestras independientes:

$$t = \frac{\bar{D}}{\sqrt{\frac{\sum D^2 - \frac{(\sum D)^2}{N}}{N(N-1)}}$$

donde

\bar{D} es la media de las diferencias de todos los pares de puntuaciones,

$\sum D^2$ es la suma de los cuadrados de las diferencias,

$(\sum D)^2$ es el cuadrado de la suma de las diferencias,

N es el número de pares de puntuaciones, y

$N - 1$ son los grados de libertad (número de pares de puntuaciones menos uno.)

Considere el siguiente ejemplo y sus pasos:

Sujetos	Puntuaciones Postest	Puntuaciones Pretest	\bar{D}	D_2
1	22	15	7	49
2	21	16	5	25
3	20	17	7	49
4	23	16	7	49
5	19	14	5	25
6	21	15	6	36
7	18	12	6	36
8	22	18	4	16
			$\sum D = 47$	$\sum D^2 = 285$

Paso 1: $\bar{D} = \frac{\sum D}{N} = \frac{47}{8} = 5.9$

Paso 2: $(\sum D)^2 = 47^2 = 2209$

Paso 3: Sustituir en la fórmula

$$\begin{aligned}t &= \frac{5.9}{\sqrt{285 - \frac{2209}{8(8-1)}}} \\&= \frac{5.9}{\sqrt{\frac{285 - 276}{56}}} \\&= \frac{5.9}{0.40} \\&= 14.75\end{aligned}$$

El estadístico t calculado (14.75) se localiza en los valores críticos de la tabla t con sus correspondientes grados de libertad ($N - 1$ o, en este ejemplo, $8 - 1 = 7$). El resultado, a partir de la tabla, es que las medias de los grupos son claramente diferentes entre sí, siendo este resultado estadísticamente significativo con una $p < 0.001$.

TABLA C.1: NÚMERO ALEATORIOS

03 47 43 73 86	36 96 47 36 61	46 98 64 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79
33 21 12 45 29	78 64 56 07 82	52 42 07 44 38	15 51 00 13 42	99 66 02 79 54
57 60 86 32 44	09 47 27 96 54	49 17 46 09 62	90 52 84 77 27	08 02 73 43 28
18 18 07 92 46	44 17 16 58 09	79 83 86 19 62	06 76 50 03 10	55 23 64 05 05
26 62 38 97 75	84 16 07 44 99	83 11 46 32 24	20 14 85 88 45	10 93 72 88 71
23 42 40 64 74	82 97 77 77 81	07 45 32 14 08	32 98 94 07 72	93 85 79 10 75
52 36 28 19 95	50 92 26 11 97	00 56 76 31 38	80 22 02 53 53	86 60 42 04 53
37 85 84 35 12	83 39 50 08 30	42 34 07 96 88	54 42 06 87 98	35 85 29 48 39
70 29 17 12 13	40 33 20 38 26	13 89 51 03 74	17 76 37 13 04	07 74 21 19 30
56 62 18 37 35	96 83 50 87 75	97 12 25 93 47	70 33 24 03 54	97 77 46 44 80
99 59 57 22 77	88 42 95 45 72	16 64 36 16 00	04 43 18 66 79	94 77 24 21 90
16 08 15 04 72	33 27 14 34 09	45 59 34 68 49	12 72 07 34 45	99 27 72 95 14
31 16 93 32 43	50 27 89 87 19	20 15 37 00 49	52 85 66 60 44	38 68 88 11 80
68 34 30 13 70	55 74 30 77 40	44 22 78 84 26	04 33 46 09 52	68 07 97 06 57
74 57 25 65 76	59 29 97 68 60	71 91 38 67 54	13 58 18 24 76	15 54 55 95 52
27 42 37 86 53	48 55 90 65 72	96 57 69 36 10	96 46 92 42 45	97 60 49 04 91
00 39 68 29 61	66 37 32 20 30	77 84 57 03 29	10 45 65 04 26	11 04 96 67 24
29 94 98 94 24	68 49 69 10 82	53 75 91 93 30	34 25 20 57 27	40 48 73 51 92
16 90 82 66 59	83 62 64 11 12	67 19 00 71 74	60 47 21 29 68	02 02 37 03 31
11 27 94 75 06	06 09 19 74 66	02 94 37 34 02	76 70 90 30 86	38 45 94 30 38
35 24 10 16 20	33 32 51 26 38	79 78 45 04 91	16 92 53 56 16	02 75 50 95 98
38 23 16 86 38	42 38 97 01 50	87 75 66 81 41	40 01 74 91 62	48 51 84 08 32
31 96 25 91 47	96 44 33 49 13	34 86 82 53 92	00 52 43 48 85	27 55 26 89 62
66 67 40 67 14	64 05 71 95 86	11 05 65 09 68	76 83 20 37 90	57 16 00 11 66
14 90 84 45 11	75 73 88 05 90	52 27 41 14 86	22 98 12 22 08	07 52 74 95 80
68 05 51 18 00	33 96 02 75 19	07 60 62 93 55	59 33 82 43 90	49 37 38 44 59
20 46 78 73 90	97 51 40 14 02	04 02 33 31 08	39 54 16 49 36	47 95 93 13 30
64 19 58 97 79	15 06 15 93 20	01 90 10 75 06	40 78 78 89 62	02 67 74 17 33
05 26 93 70 60	22 35 85 15 13	92 03 51 59 77	59 56 78 06 83	52 91 05 70 74
07 97 10 88 23	09 98 42 99 64	61 71 62 99 15	06 51 29 16 93	58 05 77 09 51
68 71 86 85 85	54 87 66 47 54	73 32 08 11 12	44 95 92 63 16	29 56 24 29 48
26 99 61 65 53	58 37 78 80 70	43 10 50 67 42	32 17 55 85 74	94 44 67 16 94
14 65 52 68 75	87 59 36 22 41	26 78 63 06 55	13 08 27 01 50	15 29 39 39 43
17 53 77 58 71	71 41 61 50 72	12 41 94 96 26	44 95 27 36 99	02 96 74 30 83
90 26 59 21 19	23 52 23 33 12	96 93 02 18 39	07 02 18 36 07	25 99 32 70 23
41 23 52 55 99	31 04 49 69 96	10 47 48 45 88	13 41 43 89 20	97 17 14 49 17
60 20 50 81 69	31 99 73 68 68	35 81 33 03 76	24 30 12 48 60	18 99 10 72 34
91 25 38 05 90	94 58 28 41 36	45 37 59 03 09	90 35 57 29 12	82 62 54 65 60
34 50 57 74 37	98 80 33 00 91	09 77 93 19 82	74 94 80 04 04	45 07 31 66 49
85 22 04 39 43	73 81 53 94 79	33 62 46 86 28	08 31 54 46 31	53 94 13 38 47
09 79 13 77 48	73 82 97 22 21	05 03 27 24 83	72 89 44 05 60	35 80 39 94 88
88 75 80 18 14	22 95 75 42 49	39 32 82 22 49	02 48 07 70 37	16 04 61 67 87
90 96 23 70 00	39 00 03 06 90	55 85 78 38 36	94 37 30 69 32	90 89 00 76 33

Fuente: Tomado de Tabla XXXII de Fisher y Yates': *Statistical Tables for Biological, Agricultural and Medical Research* (6.^a edición 1974) publicada por Longman Group UK Ltd. London (previamente publicada por Oliver and Boyd Ltd, Edinburgh) y está reproducida con permiso de los autores y editores.

TABLA C.2: VALORES CRÍTICOS PARA EL COEFICIENTE DE CORRELACIÓN DE PEARSON

df	Nivel de significación para pruebas de una cola				
	.05	.025	.01	.005	.0005
	Nivel de significación para pruebas de dos colas				
	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3223	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Fuente: Tomado de Tabla VII de Fisher y Yates: *Statistical Tables for Biological, Agricultural and Medical Research* (6.ª edición 1974) publicada por Longman Group UK Ltd. London (previamente publicada por Oliver and Boyd Ltd, Edinburgh) y está re- producida con permiso de los autores y editores.

TABLA C.3: VALORES CRÍTICOS DE CHI-CUADRADO

df	.99	.98	.95	.90	.80	.70	.50	.30	.20	.10	.05	.02	.01	.001
1	.0002	.0006	.0039	.016	.064	.15	.46	1.07	1.64	2.71	3.84	5.41	6.64	10.83
2	.02	.04	.10	.21	.45	.71	1.39	1.41	3.22	4.60	5.99	7.82	9.21	13.82
3	.12	.18	.35	.58	1.00	1.42	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	.30	.43	.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.47
5	.55	.75	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	24.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.46	23.54	26.30	29.63	32.00	39.25
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	22.62	24.77	27.59	31.00	33.41	40.79
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.80	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.78	25.04	28.41	31.41	35.02	37.57	45.32
21	8.90	9.92	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.34	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.48	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.80	35.56	38.88	42.86	45.64	54.05
27	12.88	14.12	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.45	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

Fuente: Tomado de Tabla IV de Fisher and Yates: *Statistical Tables for Biological, Agricultural and Medical Research* (6.ª edición 1974) publicada por Longman Group UK Ltd. London (previamente publicada por Oliver and Boyd Ltd, Edinburgh) y está reproducida con permiso de los autores y editores.