

determinar su confiabilidad. Imaginemos un instrumento de medición nada confiable, uno que no nos permitiese en absoluto predecir la actuación futura de los individuos, que suministrase una ordenación jerárquica de una muestra de sujetos y otra ordenación totalmente distinta en otra aplicación de la misma prueba. Con semejante instrumento no sería posible identificar ni extraer varianzas sistemáticas, puesto que las puntuaciones obtenidas serían como números tomados de una tabla de números aleatorios. Este es el caso extremo. Imaginemos ahora diversos grados de confiabilidad y desconfiabilidad en las medidas de la variable dependiente. Cuanto más confiables sean las medidas, mejor identificaremos y extraeremos las varianzas sistemáticas y menor será la varianza de error en relación con la varianza total.

Otra razón para reducir la varianza de error al máximo es permitir que las varianzas sistemáticas manifiesten su importancia, *si es que la tienen*. No podemos hacerlo si dicha varianza, y por ende el término de error, es demasiado grande. Cuando existe una relación procuramos descubrirla. Una manera de detectarla o de dejar que aparezca es encontrar las diferencias significativas que hay entre medias. Pero si la varianza de error es relativamente grande a causa de los errores incontrolados de la medición (y por lo mismo poco confiable), las varianzas sistemáticas — a las que antes dimos el nombre de "varianzas entre grupos" — no serán significativas. Así pues, aunque la relación exista, muy probablemente no será o no podrá ser descubierta.

El problema de la varianza de error se puede expresar por una sencilla fórmula matemática: recuérdese la ecuación:

$$V_t = V_b + V_e$$

donde V_t es la varianza total en un conjunto de medidas; V_b es la varianza entre grupos, que posiblemente se debe al influjo de las variables experimentales; y V_e es la varianza de error (en el análisis de varianza, la varianza entre grupos residual). Es claro que cuanto mayor sea V_e , menor deberá ser V_b , con determinada magnitud de V_t .

Más aún, examínense las siguientes ecuaciones:

$$t = \frac{\text{estadística}}{\text{error estándar de la estadística}}$$

y

$$F = \frac{V_b}{V_e}$$

Ambas expresan lo mismo: los denominadores han de ser medidas exactas del error aleatorio, a fin de que los numeradores de las fracciones de la derecha sean evaluados con precisión en busca de desviaciones notables respecto a expectativas azarosas.

Un ejemplo bien conocido ilustrará lo anterior. Recuérdese que en las discusiones sobre el análisis factorial de varianza y sobre el análisis de varianza de grupos correlacionados se hizo mención de la varianza ocasionada por la presencia de diferencias individuales en las medidas experimentales. Se dijo que si bien una buena aleatorización logra igualar a los grupos experimentales, habrá una varianza en las puntuaciones ocasionada por las diferencias individuales; por ejemplo, en las derivadas de la inteligencia, aptitud, etc. Ahora bien, en algunas situaciones, tales diferencias pueden ser muy notables. De ser así, la varianza de error y por consiguiente los denominadores de las ecuaciones t y F , serán "demasiado grandes" en comparación con los numeradores; vale decir, las diferencias individuales habrán sido distribuidas aleatoriamente entre dos, tres o cuatro grupos experimentales. A pesar de eso son fuente de varianza y por lo mismo aumentarán de manera excesiva la varianza entre grupos, la residual o la varianza típica (error estándar al cuadrado), es decir, los denominadores de las ecuaciones antes vistas.

Guías de estudio

Hemos dicho que el diseño de investigación tiene por objeto obtener respuestas a las preguntas de investigación y controlar la varianza. Explique en detalle lo que significa esta afirmación. ¿En qué forma un diseño controla la varianza? ¿Por qué debe el diseño factorial controlar más varianza que el diseño unidireccional? ¿Cómo controla la varianza un diseño que utiliza sujetos apareados o medidas repetidas de los mismos sujetos? ¿Qué relación existe entre las preguntas de investigación e hipótesis y el diseño? Al contestar las preguntas, elabore un problema de investigación para ilustrar sus aseveraciones o utilice un ejemplo tomado del texto.

Diseños inadecuados y criterios de los diseños

Todas las creaciones humanas tienen una ordenación o estructura: arquitectura, poesía, música, pintura, matemáticas, investigación científica. El ser humano concede suma importancia al contenido de sus creaciones, a menudo sin darse cuenta de que si carecen de una sólida estructura serán débiles y estériles, por muy rico e importante que sea su contenido.

Lo mismo sucede con la investigación científica. El hombre de ciencia necesita una forma viable y plástica con la cual expresar sus metas. Sin contenido — privado de una buena teoría y de buenas hipótesis y problemas — el diseño estará vacío. Pero sin forma, sin un armazón bien concebido y creado para alcanzar la finalidad que se persigue, no es posible lograr algo de valor. Por tanto, no es una exageración afirmar que muchos de los fracasos de la investigación de la conducta obedecen a la carencia de una estructura disciplinada e imaginativa.

El presente capítulo se centra sobre todo en los diseños de investigación inadecuados. Tales errores son tan comunes, especialmente en la educación, que merecen un estudio aparte, para que el estudiante pueda ser capaz de reconocerlos y comprender *por qué* son inadecuados. Este enfoque tiene una ventaja: el estudio de las deficiencias obliga a preguntar por qué algo es defectuoso, lo cual a su vez centra nuestra atención en los criterios con que se juzgan las suficiencias e insuficiencias. Por tanto, el estudio de esos diseños nos lleva a examinar los criterios del diseño de investigación. Aprovecharemos la oportunidad para describir el sistema simbólico que se empleará en este libro y para trazar una importante distinción entre la investigación experimental y la llamada investigación ex post facto.

Método experimental y método ex post facto

Una exposición sobre el diseño ha de ir precedida por una diferenciación importante: la existente entre ambos métodos. En efecto, semejante distinción es tan importante que se le dedicará un capítulo más adelante. La locución latina ex post facto significa "a partir de lo ya acontecido". Denota algo que sucede o se realiza *después* de un fenómeno que

tiene efecto retroactivo sobre el hecho. En este libro se usa por contraposición a "experimental" y se le ha dado un sentido específico y, ojalá, inequívoco.

Por experimento se entiende una investigación científica en la cual quien la realiza (el investigador) manipula o controla una o más variables independientes y observa la variable o variables dependientes en busca de una variación concomitante con la manipulación de las variables independientes. De lo dicho se deduce que el *diseño experimental* es aquél en que se *manipula* al menos una variable independiente. Hurlock manipuló los incentivos a fin de producir distintos grados de retención. Fleming y Antonen manipularon la información del cociente intelectual dada a los maestros en el estudio dedicado a los efectos de las expectativas del maestro: a algunos se les dijo que ciertos alumnos poseían cocientes intelectuales muy superiores a los que realmente tenían.

En un *verdadero* experimento, el investigador tiene el poder de asignar los sujetos a los grupos experimentales. Lo ideal es que también posea el poder de *seleccionarlos* al azar si es posible, pero por desgracia tal situación ideal a menudo se le niega. Si carece del *poder* de asignar los sujetos a los grupos experimentales o para asignar los tratamientos experimentales a los grupos, su estudio será un experimento mas no uno *verdadero*.

En la investigación ex post facto es imposible manipular y asignar los sujetos o tratamientos porque la variable o variables independientes ya ocurrieron, por decirlo así. El investigador comienza con la observación de la variable dependiente y estudia retrospectivamente las variables independientes en busca de los efectos que pudieran tener en aquella. Cuando Getzels y Jackson compararon las características de niños "no creativos" de gran inteligencia y niños "creativos" de menor inteligencia, ya se habían formado los grupos "creativos y no creativos" y los grupos muy inteligentes y menos inteligentes. (1) Dicho con otras palabras, las variables independientes ya habían "ocurrido". Getzels

(1) J. Getzels y P. Jackson, "Occupational Choice and Cognitive Functioning: Career Aspirations of Highly Intelligent and of Highly Creative Adolescents," *Journal of Abnormal and Social Psychology*, LXI (1960), 119-123.

y Jackson no podían manipular la actividad o la inteligencia, ni asignar aleatoriamente o por otro procedimiento los sujetos a los grupos ni asignarles tratamientos (a menos que hubiesen escogido otra variable independiente para manipularla, digamos los métodos de enseñanza). Aunque por lo visto la investigación experimental y la investigación ex post facto difieren notablemente en estos aspectos y en otros, comparten rasgos estructurales y de diseño que procuraremos destacar en este y en los siguientes capítulos dedicados al diseño.

El ideal científico es un experimento controlado. Con la excepción probable de la investigación taxonómica cuyo objetivo es descubrir, clasificar y medir los fenómenos naturales y los factores que los causan, el experimento controlado es el modelo óptimo de la ciencia. A no pocos estudiantes les será difícil aceptar esta afirmación bastante categórica, puesto que su lógica no se advierte de inmediato. Antes se dijo que la meta principal de la ciencia es descubrir las relaciones de los fenómenos. ¿Por qué, entonces, se concede prioridad al experimento controlado? ¿Acaso no contamos con otros métodos para descubrir las relaciones? Ciertamente que los hay. La razón fundamental de esta supremacía es que el investigador puede estar más seguro de que las relaciones descubiertas por él, son ciertas, porque las descubre en las condiciones de investigación más minuciosamente controladas que conozca el hombre. Así pues, el control es la virtud peculiar de la investigación experimental.

En conclusión, una investigación experimental realizada perfectamente, que arroja la información de que *A* se relaciona con *B*, es más confiable que una investigación ex post facto realizada de la misma manera. La razón de ello se hará evidente conforme avancemos en nuestro estudio del diseño.

Simbolismo y definiciones

Antes de adentrarnos en los diseños inadecuados, es necesario explicar el simbolismo que se utilizará en estos capítulos. *X* representa a la variable o variables independientes manipuladas experimentalmente. X_1, X_2, X_3, \dots , designan a las variables independientes 1, 2, 3 y así sucesivamente. No obstante, en general usamos una sola *X* aunque denote más de una variable independiente. (También empleamos X_1 y X_2 , etc., para designar las particiones de una variable independiente, pero la diferencia será siempre clara.) El símbolo \textcircled{X} indica que la variable independiente no se manipula, es decir, que no se encuentra bajo control directo del investigador sino que se mide o se imagina. La *Y* es la variable dependiente o, quizás más exactamente, su medida. Y_0 es la variable dependiente antes de la manipulación de *X*, y Y_1 la variable dependiente después de la manipulación. En $\sim X$, adoptamos el signo negativo de la teoría de los conjuntos, $\sim X$ ("no-*X*") significa

que la variable experimental, o sea la variable independiente *X*, no se manipula. (Nota: \textcircled{X} es una variable no manipulable y $\sim X$ es una variable manipulable que no se manipula.) Con el símbolo \textcircled{X} se indicará la asignación aleatoria de sujetos y de tratamientos experimentales a los grupos experimentales.

La explicación de $\sim X$, que acabamos de dar, no es muy precisa, pues en algunos casos ese símbolo puede significar un aspecto distinto del tratamiento de *X* y no solamente su ausencia. En la terminología antigua, el grupo experimental era aquel que recibía el llamado tratamiento experimental, *X*, en tanto que el grupo de control no lo recibía, $\sim X$. Sin embargo, este símbolo corresponde perfectamente a nuestros propósitos, en especial si comprendemos el significado generalizado de "control" que explicaremos luego. Por tanto, el grupo experimental es un grupo de sujetos que reciben algún aspecto, alguna manipulación de *X*. Al someter a prueba la hipótesis de frustración-agresión, el grupo experimental es aquél a cuyos miembros se les frustra de modo sistemático; en cambio, el grupo de control es el que no recibe tratamiento alguno.

En la moderna investigación de naturaleza multivariable es preciso extender esas nociones. No se les modifica en lo esencial, sólo se les amplía. No es raro tener más de un grupo experimental, según hemos visto. Los diferentes grados de manipulación de la variable independiente no son sólo posibles, sino que muchas veces son además deseables y hasta necesarios. Más aún, es posible tener más de un grupo de control, afirmación que al inicio se antoja un absurdo. ¿Cómo podemos tener diferentes grados de "ausencia" de tratamiento experimental? La respuesta es: porque la noción de control se generaliza. Cuando hay más de dos grupos y dos de ellos reciben un tratamiento distinto, uno o varios sirven de "grupos de control" en los demás tratamientos. Recuérdese que el control es siempre un control de varianza. Con dos o más grupos tratados en forma distinta, la varianza es producida por la manipulación experimental. Y así la noción tradicional de *X* y $\sim X$, o sea tratamiento y ausencia de tratamiento, se generaliza a X_1, X_2, \dots, X_k , es decir, las distintas formas o grados del tratamiento.

Si colocamos a la *X* dentro de un círculo, \textcircled{X} significará que el investigador "imagina" su manipulación, y supone que ha ocurrido que es la *X* de su hipótesis. Asimismo puede significar que *X* es medida y no se manipula. En realidad, estamos diciendo lo mismo en forma distinta. El contexto de la exposición hará clara la distinción. Supongamos que un sociopedagogo estudia la delincuencia y la hipótesis de frustración-agresión. Observa la delincuencia, *Y*, e imagina que sus sujetos delincuentes (designados con \textcircled{X}) fueron frustrados en sus primeros años. Todos los diseños ex post facto mostrarán este símbolo. Así pues, casi siempre \textcircled{X} denota una varia-

ble independiente que no está bajo el control del investigador.

Una observación más: cada diseño de este capítulo tendrá de ordinario una forma *a* y una forma *b*. La primera será la forma experimental, es decir, aquella en que la *X* es manipulada. La segunda será la forma ex post facto: aquella en que la *X* no se halla bajo el control del investigador, o \textcircled{X} . Evidentemente, también puede intervenir \textcircled{X} .

Diseños defectuosos

Hay tres o cuatro diseños inadecuados que se usaron y que por desgracia todavía se emplean en la investigación de la conducta. Dudamos que quienes los utilizan adviertan sus defectos. Estas son sobre todo limitaciones estructurales que ocasionan un control deficiente o nulo de las variables independientes. Los diseños defectuosos que comentaremos nos ofrecen ejemplos que, si bien de baja calidad, resultan instructivos. Imposible evitar los peligros si no se les reconoce.

Diseño 18-1: un grupo

- (a) X Y (Forma experimental)
 (b) \textcircled{X} Y (Forma ex post facto)

A este diseño se le ha denominado "estudio de casos con un solo intento" ("One-Shot Case Study"), una expresión muy apropiada.⁽²⁾ Los estudios de casos caen dentro de esta rúbrica, pero igualmente ciertos tipos de investigación. El diseño 18-1(a) es un tipo "experimental". Por ejemplo, el profesorado de una escuela decide instituir un nuevo plan de estudios y "estudiar" o "valorar" sus efectos, representados por *X* que es una variable independiente manipulada. Al cabo de un año *Y*, se estudia el aprovechamiento de los alumnos y acaso sus actitudes, pero más probablemente la opinión de los maestros. Se descubre que el aprovechamiento es-

(2) D. Campbell, "Factors Relevant to the Validity of Experiments in Social Settings," *Psychological Bulletin*, LIV (1957), 297-312. Es un excelente artículo. El autor de este libro ha sido influido por el pensamiento de Campbell y Solomon, Stouffer y Underwood, pero modificó sus formulaciones para que encajasen en la explicación que se da en esta obra del diseño. Véase a S. Stouffer, "Some Observations on Study Design," *American Journal of Sociology*, LV (1950), 355-361; R. Solomon, "An Extension of Control Group Design," *Psychological Bulletin*, XLVI (1949), 137-150; B. Underwood, *Psychological Research*. Nueva York: Appleton, 1957, caps. 4 y 5; y D. Campbell y J. Stanley, *Experimental and Quasi-Experimental Designs for Research*. Skokie, Ill.: Rand McNally, 1963. Hace poco la idea de Solomon relativa al diseño de cuatro grupos (ver el diseño 19-6 en el capítulo 19) fue aplicada a la investigación de la privación y el enriquecimiento: R. Solomon y M. Letaas, "A Control Group Design for Experimental Studies of Developmental Processes," *Psychological Bulletin*, LXX (1968), 145-150.

colar es "el mismo" o "mejor". Y así queda probado el valor del nuevo plan de estudios.⁽³⁾

El diseño 18-1(b) es la forma ex post facto del diseño con un grupo. En él se estudia o se examina la variable dependiente *Y*, y se supone o se imagina la variable independiente *X*. Un caso sería el estudio de la delincuencia y para ello se debe tomar un grupo de delincuentes juveniles y mirar hacia el pasado en busca de los factores causales que motivaron su conducta antisocial.

Desde el punto de vista científico este diseño es inútil; y lo que es peor, puede dar origen a graves errores. Como señala Campbell, el mínimo de información científica útil exige al menos una comparación formal.⁽⁴⁾ El ejemplo de la introducción de un nuevo plan de estudios requiere por lo menos la comparación entre un grupo que experimentase dicho plan y otro que no lo experimentase. El diseño carece de valor desde la perspectiva científica por que *Y*, el supuesto efecto, pudo haberse producido sin *X*, la causa probable. El supuesto efecto del nuevo plan de estudios, digamos tal o cual aprovechamiento, pudiera haber sido más o menos idéntico con cualquier programa escolar. Lo importante no es que el nuevo plan produjese o no un efecto sino que nada puede afirmarse sobre el mismo sin una comparación formal, controlada, del rendimiento de los miembros del "grupo experimental" con el rendimiento de los integrantes de algún otro grupo no experimental. Por desgracia, mucho se ha hablado sobre los supuestos resultados de semejantes "experimentos", y se han aceptado esas afirmaciones.

Hay que destacar una importante distinción. No estamos diciendo que el método sea universalmente inútil y erróneo, sino que lo es desde el punto de vista científico. Al estudiar la vida nos basamos en tales datos "experimentales". Actuamos y hablamos en base a nuestra experiencia, y ésta es la única manera en que podemos obrar. Confiamos en nuestra capacidad para usar la experiencia de manera racional y crítica. En consecuencia, el paradigma del pensamiento en las situaciones prácticas, implícito en el diseño 18-1, no está siendo criticado aunque nada se opone a ello. Los problemas se presentan cuando se le da el calificativo de científico o se piensa que lo es. Aun en las grandes empresas intelectuales es preciso utilizar este paradigma. Dentro de esta categoría caerían las brillantes observaciones y el análisis de la conducta neurótica hechos por Freud. Nuestra discusión no es con Freud, sino con la afirmación de que

(3) Con ello no se quiere decir que sea indeseable la valoración, en el sentido de una discusión y apreciación informada y minuciosa. A menudo en tales situaciones no es posible recurrir a los experimentos controlados. Pero conduce a interpretaciones erróneas llamar "investigación" a dicha estimación o establecer un experimento de este tipo.

(4) Campbell, *op. cit.*, pág. 298.

sus conclusiones fueron "establecidas científicamente".

Diseño 18-2: un grupo, con preprueba y posprueba (pretest-postest)

(a) Y_b X Y_a (Forma experimental)

(b) Y_b \textcircled{X} Y_a (Forma ex post facto)

Este diseño no es más que una versión ligeramente mejorada del 18-1. La característica esencial de este método de investigación consiste en que el grupo se compara consigo mismo. En teoría, es el mejor control puesto que todas las posibles variables independientes relacionadas con las características de los sujetos están bajo control. Vamos a describir el procedimiento expuesto en él. Un grupo es medido en la variable dependiente Y , antes de cualquier manipulación experimental. A ello se le denomina preprueba (pretest). El aprovechamiento escolar y las actitudes son dos variables dependientes que se usan comúnmente. Supongamos que se desean medir las actitudes de un grupo de sujetos hacia las mujeres. Se lleva a cabo una manipulación experimental cuya finalidad es modificarlas. El experimentador someterá al grupo a la opinión de expertos en derechos femeninos. Se interpone esa X y luego vuelven a medirse las actitudes de los sujetos. Las puntuaciones de diferencia, o sea $Y_a - Y_b$, se examinan para ver si se produjo un cambio en ellas.

A primera vista parecería un método idóneo para alcanzar la finalidad experimental. Después de todo, si las puntuaciones de diferencia poseen valor estadístico, ¿no indican un cambio de actitudes? Empero, la situación no es tan simple: hay otros factores que pueden haber contribuido a la alteración de las puntuaciones. Campbell nos ofrece una explicación excelente y detallada de ellos.(5) A continuación ofrecemos una síntesis de la misma.

Medición, historia, maduración. Primero es el efecto potencial del procedimiento de medición: al medir a los sujetos se les cambia. ¿Podría suceder que las medidas de post- X fuesen influidas, no por su manipulación, sino por un aumento de sensibilización debido al pretest? En algunas situaciones este factor no tendrá repercusiones, en otras las tendrá y considerables. Las actitudes controvertidas, por ejemplo, parecen ser especialmente susceptibles a él. A las medidas de tales variables Campbell las llama *medidas reactivas*, porque a su vez hacen

(5) *Ibid.*, págs. 298-300. El primer punto tratado, es decir, el posible efecto de la interacción producido por la preprueba, parece haber sido señalado inicialmente por Solomon en su excelente artículo citado en las páginas 140-141. Campbell profundizó en el punto original de Solomon: Campbell, *op. cit.*, págs. 298-299. Véase asimismo a Campbell y Stanley, *op. cit.*, págs. 7-12. Por ahora no nos ocuparemos de los problemas que encierra el empleo de las puntuaciones de cambio o diferencia.

reaccionar al sujeto. Las medidas de aprovechamiento, aunque probablemente sean menos reactivas, no dejan de serlo en cierto grado. Las medidas referentes a la memoria son muy reactivas. Si realiza usted una prueba ahora, lo más seguro es que más tarde recuerde cosas que se incluyeron en ella. En suma, los cambios observados quizá se deban a las medidas reactivas.

Otras dos causas importantes de varianza extraña son *historia y maduración*. Entre las pruebas efectuadas con Y_b y Y_a pueden ocurrir muchas cosas además de X . Es decir, las variables independientes extrañas pueden operar en el intervalo. Cuanto más largo sea el período, habrá mayores posibilidades de que las variables extrañas afecten a los sujetos y, en consecuencia, a las medidas de Y_a . Esto es a lo que Campbell llama *historia*. Esas variables o acontecimientos son *específicos* de una situación experimental particular. En cambio, la *maduración* abarca hechos que no son específicos de ninguna situación particular, sino que son *generales*. Están conectados con el cambio o crecimiento del organismo estudiado. La edad mental de un niño aumenta con el tiempo. Este incremento puede afectar fácilmente las medidas de aprovechamiento. Los niños aprenden durante cualquier intervalo y su aprendizaje afectará a la medida de la variable dependiente. En efecto, es uno de los rasgos exasperantes de la investigación educativa: los niños se niegan a permanecer estáticos mientras se efectúan los estudios. Por otra parte, cuanto más largo es el intervalo mayor será la posibilidad de que influyan en las medidas de la variable dependiente las variables extrañas, esas posibles e indeseables causas de varianza sistemática.

Efecto regresivo. Un fenómeno estadístico que ha despiestado e intrigado a los investigadores es el llamado *efecto regresivo*. Las puntuaciones de las pruebas cambian como un hecho estadístico de la vida: por lo regular al repetir la prueba regresan hacia la media. Dicho efecto opera a causa de la imperfecta correlación que existe entre las puntuaciones de la preprueba y las de la posprueba: Si $r_{ab} = 1.00$, no se producirá el efecto; si $r_{ab} = .00$, el efecto será máximo ya que la media es la mejor predicción de cualquier puntuación de posprueba hecha a partir de la puntuación de preprueba. Con las correlaciones moderadas e importantes, pero imperfectas, que se encuentran en la práctica el efecto neto es que las puntuaciones inferiores obtenidas en la preprueba tienden a ser superiores y las superiores tienden a ser inferiores en la posprueba, cuando de hecho no se ha operado ningún cambio en la variable dependiente. Por consiguiente, si los sujetos que obtienen baja puntuación —los pobres, por ejemplo— intervienen en un estudio, sus puntuaciones en la posprueba probablemente sean mayores que las obtenidas en la preprueba, en virtud del efecto regresivo. Esto puede engañar al investigador y hacerle creer que su parti-

cipación experimental fue eficaz cuando en realidad no fue así. De manera análoga, si se estudian exclusivamente los que obtienen elevadas calificaciones en la preprueba podríamos concluir erróneamente, por ejemplo, que un programa escolar ha ejercido sobre ellos un efecto deprimente. No sucede necesariamente así. Las puntuaciones bajas pueden obedecer al efecto de regresión.

¿Cómo funciona esto? En todo conjunto de puntuaciones intervienen numerosos factores fortuitos. (6) En la preprueba, la casualidad hace que algunas puntuaciones elevadas sean mayores de lo que "debieran ser". Lo mismo sucede con las bajas. En la posprueba es poco probable que se conserven las primeras porque los factores que las hicieron elevadas fueron fortuitos y en definitiva no están correlacionados en la preprueba y posprueba. De ahí que quien obtiene puntuaciones elevadas tienda a perder terreno en la posprueba. Un razonamiento semejante se le aplica a los que logran puntuaciones bajas.

Cualquier diseño de investigación ha de construirse teniendo presente el efecto de regresión. El diseño 18-2 no ofrece ninguna forma de controlarlo. Si incluye un grupo de control se podría "controlar" dicho efecto, puesto que ambos grupos serían sometidos a una preprueba y posprueba. Si la manipulación experimental ha producido un efecto "real", éste debe manifestarse con creces sobre el efecto regresivo; vale decir, en igualdad de circunstancias las situaciones de ambos grupos son afectadas de igual manera por el efecto regresivo y por otros efectos. En consecuencia, si los grupos difieren en la posprueba se deberá a manipulación experimental.

El diseño 18-2 es insuficiente no tanto porque operen las variables extrañas y el efecto regresivo (aquellas pueden intervenir siempre que haya un intervalo entre la preprueba y la posprueba) sino porque ignoramos si han obrado o no, si han afectado o no a las medidas de la variable dependiente. No ofrece la posibilidad de verificar dichas influencias.

Diseño 18-3: con simulacro de preprueba y posprueba

	X	Y_a
Y_b	—	—

(6) Gran parte de nuestra explicación está tomada de la lúcida y excelente exposición de Anastasi sobre el efecto regresivo: A. Anastasi, *Individual Differences*, 3a. edición, Nueva York: Macmillan, 1958, págs. 203-205. Se encontrará una exposición más completa e igualmente buena en R. Thorndike, *Concepts of Over and Underachievement*. Nueva York: Teachers College Press, 1963, págs. 11-15. Este pequeño libro constituye una demostración de generalidad analítica, que deben examinar minuciosamente quienes se dedican a la investigación. Con claridad y autoridad describe Thorndike algunos de los arduos problemas de la investigación; sugiere además soluciones a los problemas.

El extraño nombre de este diseño deriva en parte de su misma naturaleza. Igual que el 18-2, es un diseño con preprueba y posprueba. Pero en vez de recurrir a las medidas de una y otra, se emplean como medidas de preprueba las de otro grupo lo más aproximadas al grupo experimental y, por tanto, funciona como una especie de grupo de control (la línea entre los dos niveles indica grupos separados). Este diseño llena el requisito de tener un grupo de control y constituye un gesto de condescendencia hacia la comparación exigida por la investigación científica. Sin embargo, es lamentable que los controles sean demasiado débiles, por nuestra incapacidad de saber si ambos grupos eran equivalentes antes de introducir X : la manipulación experimental.

Diseño 18-4: dos grupos, sin control

(a) $\frac{X}{\sim X} \quad \frac{Y}{Y}$ (Forma experimental)

(b) $\frac{\textcircled{X}}{\sim \textcircled{X}} \quad \frac{Y}{\sim Y}$ (Forma ex post facto)

Este diseño es común. En (a) el grupo experimental es sometido al tratamiento X ; el grupo de control que se considera o se supone similar al grupo experimental no recibe ese tratamiento. Las medidas de Y se comparan a fin de averiguar el efecto de X . Los grupos o sujetos se toman "tal como son", o se aparean. La versión ex post facto del mismo diseño aparece encerrada en un círculo (b). Se observa que ocurre un efecto Y en un grupo (línea superior) pero no en el otro, o bien que se presenta en el otro pero en menor grado (indicado por $\sim Y$ en la línea inferior). Se descubre que el primer grupo ha experimentado X y que el segundo no.

El diseño adolece de una deficiencia básica. Se supone que ambos grupos son iguales en las variables independientes que no sean X . A veces es posible verificar aproximadamente la igualdad si se les compara con distintas variables pertinentes, tales como sexo, edad, ingreso económico, inteligencia, capacidad. Conviene hacerlo siempre que sea posible; pero como dice Stouffer: "... con demasiada frecuencia hay un portón de par en par por donde pueden pasar otras variables incontroladas". (7) Puesto que la aleatorización no se usa —es decir, los sujetos no son asignados al azar a los grupos— así que la suposición de que los grupos son iguales es incorrecta. El diseño 18-4(b) es más débil que el (a), en éste se manipula a X , no así en aquél. No obstante, ambas versiones adolecen de una grave falta de control de las variables independientes, ya que no existe "aleatorización".

(7) Stouffer, *op. cit.*, pág. 522.

Criterios del diseño de investigación

Después de examinar algunas de las principales deficiencias de los diseños inadecuados, tenemos los elementos para exponer lo que podría llamarse criterios para el diseño de investigación. Junto con ellos enunciaremos ciertos principios que deben guiar al investigador. Por último, los criterios y principios se relacionarán con las nociones de Campbell sobre la validez interna y externa que, en cierto modo, expresan los criterios en forma distinta.

¿Contesta el diseño

las preguntas de la investigación?

El criterio esencial de un diseño puede expresarse en una pregunta: ¿contesta el diseño a las preguntas planteadas? o ¿somete a prueba las hipótesis en forma adecuada? Acaso la falla más grave de los diseños propuestos frecuentemente por los estudiantes es que no sirven para responder debidamente a las preguntas de la investigación. Un ejemplo corriente de incongruencia entre las preguntas y la hipótesis, por una parte, y el diseño, por la otra, es igualar a los sujetos en base a motivos ajenos a la investigación y luego utilizar un diseño de tipo grupo experimental-grupo de control. Los estudiantes suponen a menudo que por el simple hecho de aparear a los alumnos digamos en inteligencia y en sexo, sus grupos experimentales son idénticos. Han oído que es preciso el apareamiento de los sujetos para alcanzar "control" y que hay que tener un grupo experimental y un grupo de control. No obstante, muchas veces las variables de igualdad son ajenas a las finalidades de la investigación. Es decir, si no hubiese relación alguna entre el sexo y la variable dependiente por ejemplo, entonces el apareamiento de los sujetos en cuanto al sexo carecería de importancia.

Otro ejemplo de esta deficiencia es el caso en que se necesitan tres o cuatro grupos experimentales —por ejemplo, el tratamiento experimental de tres grupos y uno de control o cuatro grupos con diferentes magnitudes de X —, sin embargo, el investigador no emplea sino dos porque se le ha dicho que un grupo experimental y un grupo de control son indispensables y recomendables.

Otro ejemplo es el que se comentó en el capítulo 17 sobre la verificación de una hipótesis de interacción, lo cual se logra realizando dos experimentos separados. He aquí la hipótesis que se quería demostrar. La discriminación en las admisiones de los colegios era una función tanto del sexo como del nivel de capacidad, a saber: se discrimina a un mayor número de mujeres con poca habilidad en comparación con los varones de la misma habilidad. Se trata de una hipótesis de interacción y probablemente exige un diseño de tipo factorial. Establecer dos experimentos (uno

para los candidatos de alta habilidad y otro para los de baja) es una práctica deficiente porque por ese diseño, según vimos, no se puede demostrar en forma terminante la hipótesis. Análogamente, aparear a los sujetos en cuanto a capacidad y luego implantar un diseño bigrupal sería desviarse por completo de la pregunta de investigación. Tales consideraciones nos llevan a una regla general y obvia.

Diseña la investigación para dar respuesta a las preguntas que la motivaron.

Control de las variables independientes extrañas

El segundo criterio es el control, que significa control de las variables independientes: tanto las del propio estudio de investigación como las extrañas. Por supuesto, estas últimas influyen en la variable dependiente pero no forman parte del estudio. Por ejemplo, en la investigación del capítulo 17 relativa a las universidades que practicaban la discriminación en las admisiones de hombres y mujeres, la ubicación geográfica de las universidades puede ser una variable extraña, capaz de oscurecer los resultados. De este modo, si las universidades del este excluyeran más mujeres que las del oeste, la ubicación geográfica podría ser una fuente extraña de varianza en las medidas de las admisiones, por lo que debe ser controlada de alguna manera. El criterio denota asimismo el control de las variables del estudio. No ahondaremos ahora en este problema, puesto que ya ha sido explicado y se seguirá tratando. Pero cabe formular una pregunta: ¿controla debidamente este diseño las variables independientes?

La mejor manera de dar una respuesta satisfactoria está expresada en el siguiente principio:

Practique la asignación aleatoria siempre que sea posible: seleccione al azar a los sujetos; asigne los aleatoriamente a los grupos y asigne aleatoriamente los tratamientos experimentales a los grupos.

Aunque no sea posible seleccionar de manera aleatoria a los sujetos, tal vez puedan ser asignados a los grupos mediante ese procedimiento. Con ello se "aparean" los grupos en el sentido estadístico que se comentó en las partes 4 y 5. Si tampoco eso es factible, debe procurarse, por todos los medios, hacer una asignación aleatoria de los tratamientos experimentales a los grupos. Y si dichos tratamientos son aplicados en momentos distintos por diversos experimentadores, ambos han de emplear el método aleatorio. En la medida en que la "aleatorización" se ignore o se haga imposible, será débil el diseño de investigación, pues en ella descansa la estructura entera del razonamiento probabilístico-estadístico.

De modo que volvemos a repetir: *practique* la asignación aleatoria siempre que sea posible.

El principio que hace pertinente la aleatorización es complejo y difícil de poner en práctica:

Controle las variables independientes para reducir al mínimo la posibilidad de que intervengan las causas extrañas e indeseables de varianza sistemática.

Hemos visto antes que, en teoría, la "aleatorización" cumple con dicho principio (ver cap. 8). Cuando comprobamos la validez empírica de una proposición condicional ("si p ocurre, se realizará también q "), manipulamos p y observamos que q covaría con la manipulación de p , ¿pero qué tan seguros podemos estar de que nuestra proposición condicional es verdaderamente "cierta"? Nuestra confianza depende directamente de lo completo y adecuado de los controles. Si recurrimos a un diseño similar a los diseños del 18-1 al 18-4 será imposible tener mucha confianza en la validez empírica de la proposición condicional, puesto que nuestro control de las variables independientes extrañas es débil o nulo. Dado que semejante control no siempre se logra en numerosas investigaciones de orden psicológico, sociológico y educacional, ¿hemos de renunciar por completo a él? Claro que no. Pero debemos advertir las debilidades de un diseño intrínsecamente defectuoso y deficiente. Con frecuencia es posible construir y utilizar un diseño mejor del que inicialmente se consideraba factible.

Generalizabilidad

El tercer criterio, *generalizabilidad*, no está subordinado a los otros porque pertenece a una clase diferente. Este es un punto importante que muy pronto se tomará claro. Simplemente quiere decir: ¿podemos generalizar los resultados de un estudio a otros sujetos, otros grupos y otras condiciones? He aquí otra forma quizá más precisa de formular la interrogante: ¿cuánto podemos generalizar los resultados del estudio? o ¿a quién y a qué cosa podemos generalizarlos? Se trata probablemente de la pregunta más compleja y difícil que se puede hacer en base a los datos de la investigación, pues no sólo alude a cuestiones técnicas como el muestreo y el diseño de investigación, sino también a problemas más vastos de la investigación básica y aplicada. Por ejemplo, en la investigación básica, la generalizabilidad no es la consideración esencial, porque el interés central son las relaciones de las variables y la causa por las que se relacionan de una determinada forma. (8) Se hace hincapié más en los aspectos in-

(8) Se puede hallar una breve explicación de la investigación básica y aplicada en F. Kerringer, "Research in Education." En R. Ebel, V. Noll y R. Buser, Eds., *Encyclopedia of Educational Research*, 4a. edición, Nueva York: Macmillan, 1969, págs. 1127-1143, en especial la página 1128. En el artículo se citan varias obras de consulta dedicadas a la investigación básica y aplicada.

ternos del estudio que en los externos. En cambio, en la investigación aplicada el interés central demanda mayor preocupación por la generalizabilidad, pues se desea extender los resultados a otras personas y situaciones. Si el lector reflexiona sobre los dos ejemplos de investigación básica y aplicada que se citan a continuación, comprenderá mejor esta distinción.

En el capítulo 14 examinamos un estudio de Berkowitz sobre la suscitación de hostilidad, sobre el antisemitismo y la agresión desplazada. Salta a la vista que se trata de una investigación básica: el interés central se enfocó a la relación que existe entre hostilidad, antisemitismo y agresión desplazada. Nadie es lo bastante necio para afirmar que a Berkowitz no le interesaba la hostilidad, el antisemitismo, ni la agresión desplazada, en general; sin embargo, se daba prioridad a las relaciones de las variables del estudio. Confróntese el referido estudio con el de Walster y otros para saber si las universidades discriminan a la mujer. Claro está que Walster y sus colegas fueron muy explícitos respecto a los aspectos internos de su trabajo; pero pudieron tener otro interés: ¿practican las universidades la discriminación en términos generales? Su estudio es sin duda una investigación aplicada, si bien no podemos afirmar que falta interés por la investigación básica. Las consideraciones de la siguiente sección nos ayudarán a clarificar lo que es la generalizabilidad.

Validez interna y externa

Dos criterios generales del diseño fueron estudiados a fondo por Campbell y por Campbell y Stanley. (9) Sus nociones constituyen una de las más notables, importantes e iluminadoras aportaciones a la metodología de la investigación en las últimas dos o tres décadas.

La *validez interna* plantea la pregunta: ¿causó la manipulación experimental una diferencia realmente significativa? Los tres criterios del capítulo anterior son en realidad aspectos de validez interna. En efecto, cualquier cosa que afecte a los controles de un diseño se convierte en un problema de validez interna. Por ejemplo, si en un diseño se tiene poca o ninguna confianza en las relaciones mostradas por las diferencias entre los grupos experimentales surgirá un problema de validez interna. Por tanto, cuando en adelante se emplee esa expresión será una forma general de designar el criterio de control comentado antes.

La *validez externa*, que es un criterio difícil de llenar, significa *representatividad* o *generalizabilidad*. Cuando un experimento ha finalizado y se ha encontrado una relación, ¿a qué poblaciones se pue-

(9) Campbell, *op. cit.*, Campbell y Stanley, *op. cit.* Instamos a los lectores a que estudien esas fuentes porque nuestra exposición se limita a definir y destacar la validez interna y externa.

de generalizar? ¿Podemos afirmar que *A* se relaciona con *B* en todas las escuelas? ¿En todos los alumnos de sexto grado? ¿En todos los alumnos de sexto grado dentro de este sistema escolar o únicamente en los que asisten a esta escuela? ¿O han de limitarse los hallazgos a los de octavo grado con quienes trabajamos? Es una cuestión científica de suma importancia que siempre debe ser formulada —y contestada.

No sólo debe ponerse en duda la generalizabilidad de la muestra. Es preciso además preguntar sobre la representatividad de los estudios desde el punto de vista de la ecología y de las variables. Si se modifica el ambiente social donde se llevó a cabo el experimento, ¿permanecerá inalterada la relación de *A* y *B*? ¿Habrá relación entre *A* y *B* si el estudio se repite en una escuela de clase social baja? ¿En una escuela del oeste? ¿En una escuela del sur? Tales interrogantes se refieren a la *representatividad ecológica*.

La *representatividad de las variables* es más sutil. Una pregunta que muchas veces se omite pero que conviene hacer, es: ¿son representativas las variables de la investigación? Cuando un investigador trabaja con variables psicológicas y sociológicas, da por sentado que sus variables son "constantes". Si observa una diferencia de rendimiento entre niños y niñas, supone que el sexo es una variable "constante".

En el caso de variables tales como el aprovechamiento, agresión, aptitud y ansiedad, ¿puede el investigador suponer que la "agresión" de sujetos suburbanos es la misma que la que se manifiesta en los barrios bajos? ¿Es igual la variable en un suburbio de Europa? La representatividad de la "angustia" es más difícil de determinar. Al hablar de "angustia", ¿a qué clase nos referimos? ¿Son todas idénticas? Si la angustia es manipulada en una situación mediante instrucciones verbales y en otra por medio de una descarga eléctrica, ¿son idénticas las dos angustias provocadas? Por ejemplo, si se le manipula mediante la instrucción experimental, ¿será idéntica a aquella que se mide por una escala? (10) De lo dicho se deduce que la representatividad de la variable es otro aspecto de un problema más amplio de validez externa y por lo mismo de generalizabilidad.

Salvo que se tomen precauciones especiales y se hagan esfuerzos particulares, los resultados de la investigación por lo general no son representativos y, por ende, no son generalizables. Campbell y Stanley sostienen que la validez interna es la condición *sine qua non* del diseño, pero que la validez del diseño ideal ha de ser fuerte tanto interna como externamente, aunque a veces ambas sean contra-

ditorias. (11) Este punto es evidente. En los siguientes capítulos se insistirá en la validez interna, sin dejar de atender la validez externa.

En este capítulo se partió de un enfoque negativo, en la creencia de que el desenmascaramiento de procedimientos defectuosos pero de uso común, junto con la exposición de sus principales debilidades, constituiría un buen punto de partida para estudiar el diseño de investigación. Nuestra manera de proceder está plenamente justificada porque en capítulos precedentes el lector se encontró con diseños adecuados y con los principios que los rigen, aunque no se les haya calificado así. Hay otros diseños inadecuados, pero lo son atendiendo sólo a los principios que rigen su estructura. Cabe subrayar eso porque en el siguiente capítulo encontraremos que un diseño bien estructurado puede utilizarse de modo indebido. Por ello es preciso conocer y comprender las dos causas de deficiencia en la investigación: diseños intrínsecamente defectuosos y diseños intrínsecamente correctos que se emplean en forma incorrecta.

Guías de estudio

1. El cuerpo docente de una universidad recibe el cargo de estudiar el programa de un plan de estudios implantado el año anterior. Todos los estudiantes de primer grado tendrán que seguir dicho programa. Explique los problemas y dificultades que enfrentarán los investigadores. ¿Qué tan confiables serán sus hallazgos?
2. El personal docente de una universidad pedagógica decide iniciar un nuevo plan de estudios, destinado a los alumnos no graduados, y averiguar su eficacia. A un grupo de investigación del profesorado se le pide que estudie el programa durante dos años. Se desea disponer de un grupo con el cual se va a comparar el grupo que seguirá el nuevo plan de estudios, y por ello solicita que el programa ordinario siga vigente dos años y que se permita a los alumnos ofrecerse como voluntarios para el nuevo o el viejo programa. Los investigadores piensan que así se contará con un grupo experimental y con un grupo testigo.
Examine críticamente la propuesta mencionada. ¿Tendría usted confianza en sus hallazgos una vez transcurridos los dos años? Dé las razones que lo impulsan a reaccionar en forma positiva o negativa.
3. Imagine que usted es profesor en una escuela para graduados y que se le invita a juzgar el valor de una tesis doctoral. El que la presenta es un inspector que está ensayando un nuevo tipo de administración en su sistema escolar. Pro-

yecta estudiar los efectos de la nueva administración durante un periodo de tres años y luego escribir su tesis. Para garantizar la objetividad de su trabajo no estudiará ninguna otra situación escolar.

Discuta la propuesta del candidato, teniendo presente esta pregunta: si termina el estudio, ¿se le debe otorgar el título académico? ¿Es conveniente su estudio para un trabajo doctoral?

4. En su opinión, ¿debe toda investigación seguir estrictamente el criterio de generalizabilidad? Si es así, ¿por qué? En caso contrario, ¿por qué no? ¿En qué campo hay mayor cantidad de investigaciones básicas: en psicología en educación? ¿Por qué? ¿Qué consecuencias tiene su conclusión para la generalizabilidad?
5. ¿Qué relación tiene la reproducción de una investigación con la *generalizabilidad*? Explique su respuesta.

(10) Se refiere al problema de la equivalencia de definiciones. Véase A. Baldwin, "The Study of Child Behavior and Development." En P. Mussen, ed., *Handbook of Research Methods in Child Development*. Nueva York: Wiley, 1960, págs. 12-13.

(11) Campbell y Stanley, *op. cit.*