



UNIVERSIDAD BÍBLICA
LATINOAMERICANA
PENSAR • CREAR • ACTUAR

BACHILLERATO EN CIENCIAS TEOLÓGICAS

LECTURA SESIÓN 7

CT 114 PASTORALES ESPECÍFICAS

Alvira, Francisco. “Diseños de investigación social: criterios operativos”. En *El análisis de la realidad social. Métodos y técnicas de investigación*, compilado por Manuel García Ferrando, Jesús Ibáñez y Francisco Alvira, 99-125. Madrid: Alianza, 2016.

Reproducido con fines educativos únicamente, según el Decreto 37417-JP del 2008 con fecha del 1 de noviembre del 2012 y publicado en La Gaceta el 4 de febrero del 2013, en el que se agrega el Art 35-Bis a la Ley de Derechos de Autor y Derechos Conexos, No. 6683.

3. Diseños de investigación social: criterios operativos

Francisco Alvira Martín

La delimitación de *objetivos* de una investigación, así como de las *hipótesis* (en su caso) y definición de las *variables* pertinentes, es una tarea ardua, difícilmente estructurable, que queda más para la imaginación del investigador que para sus conocimientos técnicos. Jesús Ibáñez, en el capítulo anterior, ha tratado en cierta medida de estos aspectos esenciales de toda investigación, aspectos que constituyen además un prerrequisito para la misma. El recurso a la literatura científica existente sobre el tema ayuda bastante en dicha delimitación y definición así como el estudio de ejemplos aislados y la realización de investigaciones exploratorias. Los criterios para lograrlo hoy por hoy están poco estudiados y formalizados.

Una vez especificados los objetivos, definidas las hipótesis y variables, el investigador necesita elaborar un *plan, proyecto o diseño* que le guíe en el proceso de recoger, analizar e interpretar las observaciones/datos que lleve a cabo. Precisamente un diseño de investigación se define como el plan global de investigación que integra de un modo coherente y adecuadamente correcto técnicas de recogida de datos a utilizar, análisis previstos y objetivos. Dicho de otra manera el diseño de una investigación intenta dar de una manera clara y no ambigua respuestas a las preguntas planteadas en la misma.

De aquí se deduce que un buen diseño de investigación es, ante todo, función de los objetivos fijados y que no puede hablarse de un solo tipo de diseño de investigación como diseño prototípico a utilizar en todo momento y lugar. Así la experimentación puede constituir un buen diseño para responder a pre-

guntas de tipo casual, pero evidentemente tiene escaso valor para responder a pregunta de tipo descriptivo.

Si el diseño depende de los objetivos, el lector podría pensar que resulta imposible hablar de diseños de un modo abstracto. Sólo conociendo los objetivos específicos de una investigación cabría construir un diseño adecuado. Pero los objetivos de cualquier investigación pueden clasificarse siempre como:

- descriptivos y/o
- explicativos¹.

Desde un punto de vista filosófico, distinguir entre descripción y explicación no resulta sencillo, pero sí lo es desde un criterio de pragmatismo. Querer saber cuántas personas tienen intención de votar a un partido político es un objetivo descriptivo; querer averiguar por qué se vota o no a determinados partidos o por qué determinadas personas votan y otras se abstienen es un objetivo explicativo.

La mayoría de las investigaciones que se hacen en sociología son esencialmente descriptivas, aunque muchas veces intentan también describir relaciones entre variables o descienden a una descripción tan minuciosa que parece acercarse a la explicación. A pesar de esto, la dicotomía explicación/descripción en diseños de investigación es relevante y básica.

Las variables de una investigación

La delimitación de las hipótesis y la definición de las variables que serán utilizadas en una investigación son pasos necesarios que deben ser dados antes de plantearse la elaboración de un diseño de investigación. No obstante, sean cuales sean las variables a utilizar por el investigador, éstas pueden clasificarse en cuatro categorías:

1. *Variables explicativas*. Son las variables que vamos a utilizar en la investigación de un modo consciente, es decir, son aquéllas que pretendemos medir o recoger. De acuerdo con nuestros objetivos o hipótesis determinamos qué características de nuestras unidades de análisis queremos medir. Estas características que pretendemos medir —y que al llevar a cabo la investigación mediremos— constituyen las variables explicativas.

A veces se distingue en las investigaciones explicativas entre variables explicativas y explicadas, predictivas o «predichas», variables independientes y variables dependientes. Pues bien, todas ellas entran dentro de esta categoría de variables explicativas entendida como el conjunto de características de las unidades de análisis que queremos medir o sobre las que queremos recoger información. Se corresponden con los objetivos de la investigación y están directamente determinadas por éstos.

3. Diseños de investigación social: criterios operativos

2. *Variables controladas.* Son aquellas variables o fuentes de variación que el investigador controla a través del diseño de investigación escogido y lo hace bien en el momento en que se lleva a cabo la investigación mediante el propio diseño (control *a priori*), bien en el análisis de los datos (control *a posteriori*). Muchas veces el control se lleva a cabo en ambos momentos, es decir, tanto en el diseño como en el análisis.

Si el control se realiza *a posteriori*, en el proceso de estimación y análisis, es necesario que dichas variables hayan sido previamente medidas por el investigador. En ese caso no todas las variables utilizadas en la investigación serían variables explicativas en el sentido que se ha dado a este término en el apartado anterior, sino que parte serían variables controladas y parte variables explicativas.

3. *Variables perturbadoras.* Son aquellas variables que pueden confundirse con las variables explicativas al no haber sido controladas por el investigador.

Una relación entre dos variables siempre admite la posibilidad de que exista una explicación alternativa (una tercera variable que haga espúrea la anterior relación); pues bien, esta posibilidad de explicaciones alternativas —o de terceras variables— es la que queda incluida en la categoría de variables perturbadoras. Estas terceras o cuartas variables constituirán variables perturbadoras siempre que no se hayan controlado.

Cualquier diseño de investigación intenta precisamente hacer frente a la posible presencia de variables perturbadoras tratando de lograr el control del máximo posible de éstas, es decir, tratando de lograr que las variables perturbadoras se transformen en controladas.

4. *Variables aleatorias o estocásticas.* Son variables que no son controladas, pero que no introducen sesgo en la investigación al resultar simplemente en errores aleatorios o de azar, es decir, error no sistemático. Pueden verse como variables que en principio ejercen una influencia determinada sobre las variables explicativas, pero que es muy pequeña y al ser muy numerosas éstas se traduce en errores de carácter aleatorio.

Criterios para evaluar diferentes diseños de investigación

Existen muchas maneras de evaluar un diseño; tantas como criterios/dimensiones tienen éstos. Así, puede evaluarse un diseño de investigación desde el punto de vista de la adecuación/coherencia entre los recursos disponibles (humanos y financieros) y los recursos que previsiblemente serán necesarios; o puede evaluarse en lo que respecta a la posibilidad práctica de realizarlo en el tiempo previsto o incluso a la posibilidad/imposibilidad global de llevarlo a cabo.

Aunque estos criterios/dimensiones de evaluación de diseños son importantes e imprescindibles, no son los más importantes desde un punto de vista

metodológico/sustantivo. La evaluación debe hacerse centrándose en el análisis de la adecuación entre *diseño y objetivos* perseguidos. Naturalmente esto presupondría un listado de posibles objetivos; pero no es necesario si seguimos manteniendo la dicotomía básica de objetivos *explicativos/objetivos descriptivos*.

En 1967 Campbell y Stanley, en un memorable artículo, acuñaron dos criterios valorativos originales:

- el criterio de validez interna
- el criterio de validez externa.

A los que se han ido añadiendo en estos últimos años otros dos criterios más:

- el criterio de validez de constructo, y
- el criterio de validez estadística.

Estos cuatro criterios permiten una aproximación fructífera a la evaluación de diseños de investigación, aunque, como se verá más adelante, no son los únicos criterios existentes.

El criterio de validez interna

Existen tres criterios que normalmente se utilizan para determinar relaciones de tipo causal entre variables:

- a) La existencia de *covariación* entre las variables implicadas, es decir, que X e Y varíen conjuntamente.
- b) Antecedencia temporal por parte de lo que consideremos «causa» o variable independiente.
- c) Inexistencia de alternativas explicativas plausibles de la variable que tomemos como dependiente, o dicho de otra manera, que la relación entre X e Y sea auténtica/real y no espúrea.

Este último criterio/condición es el más difícil de conseguir, puesto que para cualquier relación bivariada cabe pensar en terceras variables que pudieran hacer espúrea dicha relación.

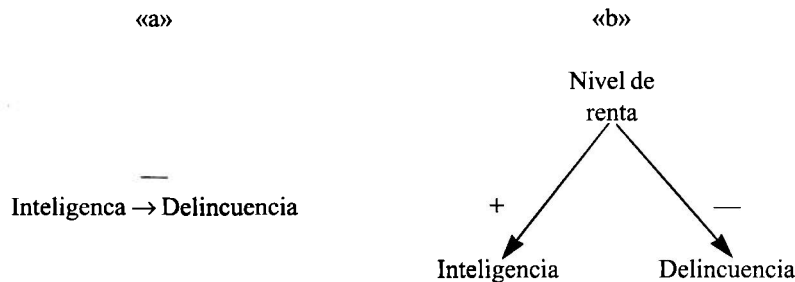
Ejemplos clásicos de relaciones espúreas abundan en todos los manuales, así la relación entre daños producidos en incendios y número de bomberos, la del número de cigüeñas y tasa de nacimientos o la de fumar e interesarse por la política. No es que el daño producido en un incendio dependa del número de bomberos que van a apagarlo, sino que ambos dependen de la magnitud del incendio; cuanto mayor es éste, normalmente mayores daños se pro-

3. Diseños de investigación social: criterios operativos

ducirán y además mayor número de bomberos irán a apagarlo. De la misma manera, no es que en las zonas con mayor número de cigüeñas haya una mayor natalidad, sino que ambas cosas se dan en zonas rurales y no es que las mujeres que fuman se interesen más por la política, sino que las mujeres con mayores ingresos a la vez fuman más y se interesan más por la política.

Pues bien, se dice que un diseño de investigación tiene validez interna cuando *todas las posibles explicaciones alternativas quedan controladas* por el diseño, de tal modo que los resultados de la investigación correspondiente tendrán una interpretación inequívoca; bien la relación o relaciones investigadas se comprueban, bien se rechazan. En términos de la anterior tipología de variables, se dice que un diseño tiene validez interna cuando no existen variables perturbadoras, estando éstas controladas o aleatorizadas. Conviene señalar qué significa y qué no significa la validez interna:

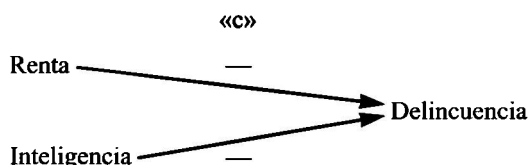
1. Por ejemplo, si nosotros hemos llevado a cabo una investigación y nuestros datos muestran una relación entre dos variables, para averiguar si existe validez interna o no tendremos que preguntarnos si hay posibles explicaciones alternativas que harían espúrea dicha relación y si estas explicaciones alternativas están o no controladas. Por ejemplo, existen suficientes datos que apuntan la posibilidad de que inteligencia y delincuencia estén relacionadas en el sentido de que los delincuentes tienen un menor nivel de inteligencia. En las investigaciones que muestran esta relación habría que analizar si la explicación de la renta como causa común o del nivel de educación o del desempleo están o no controladas como alternativas explicativas. Probablemente no es la menor inteligencia la que lleva a la delincuencia, sino que situaciones económicas negativas (bajo nivel de renta) van asociados a niveles de inteligencia más bajos y también a una mayor delincuencia. La relación no sería como se esquematiza en el gráfico «a», sino tal y como aparece en el gráfico «b».



Lo importante no es la presencia de covariaciones en los resultados de una investigación, sino la presencia o ausencia de alternativas explicativas, o dicho de otro modo, el control de las explicaciones alternativas, control que como ya se dijo se puede realizar *a priori* en el propio diseño o

a posteriori en el análisis. Al hablar de validez interna de un diseño me estoy refiriendo ante todo al control *a priori*, al control introducido en el propio diseño.

2. La validez interna no es un problema que se derive de la existencia de más de una posible explicación a un fenómeno o a una variable dependiente, sino que dicha posible explicación tiene que referirse a la idea de causa o antecedente común. En el ejemplo anterior podrá argumentarse que tanto la inteligencia como la renta influyen en la delincuencia y discutir sobre la relativa importancia de cada una de las dos variables independientes dentro de un esquema como el que aparece en el gráfico «c».



Pero el que además de la inteligencia pueda haber otras variables como la renta o el nivel de educación que tengan un impacto sobre la delincuencia no afecta para nada, siempre que el esquema «c» fuera el verdadero, a la relación entre inteligencia y delincuencia, por lo que una discusión de este tipo no versaría sobre la validez interna. Sólo si existe una explicación alternativa no controlada que haga espúrea una relación objeto de nuestra investigación estaremos hablando de validez interna.

3. Por último, conviene señalar que la validez interna debe verse más que como todo o nada como un ideal al que hay que aproximarse y que se alcanza en menor o mayor grado. Un diseño de investigación tiene mayor —o menor— validez interna que otro, pero casi nunca cabe hablar de que un diseño tiene validez interna y otro no la tiene. El grado de validez interna que un diseño tiene es lo que limita y constriñe la interpretación analítica de los resultados de una investigación; esta interpretación será más o menos adecuada, más o menos «verídica» según el grado de validez interna que posea el diseño.

El primer paso para poder controlar posibles hipótesis explicativas alternativas o dicho de otra manera, variables perturbadoras, es conocer y ser consciente de cuáles puedan ser éstas. El número de posibles hipótesis explicativas alternativas es casi ilimitado dependiendo de los objetivos de la investigación y de las propias hipótesis que la misma quiera contrastar. La tarea parece, por tanto, ardua e imposible. Sin embargo, es posible codificar en categorías genéricas las fuentes/orígenes de posibles explicaciones alternativas. Esta es precisamente la labor desarrollada por Campbell y Stanley en su artículo original de 1967. La lista original de ambos autores comprende nueve categorías que pueden resumirse en cuatro a efectos expositivos (véase Sellitz y otros, 181 —cap. 5—):

3. Diseños de investigación social: criterios operativos

1. Características iniciales y/o experiencia previa de las unidades de análisis.
2. Acontecimientos externos a la investigación que tengan lugar durante el desarrollo de la misma.
3. El proceso de maduración de los sujetos de la investigación.
4. Los propios efectos reactivos de la utilización de diferentes técnicas de investigación, en especial técnicas de medición.

Puesto que la referencia a un diseño de investigación específico facilita la comprensión de esta agrupación de categorías, fuente potencial de explicaciones alternativas, supongamos que estamos interesados en medir el efecto de una campaña política sobre la conducta electoral de los votantes y que hemos previsto utilizar un panel, es decir, una medición antes de la campaña y otra después de la misma con el mismo grupo de personas. Supongamos, asimismo, que pensáramos medir el impacto de la campaña por la diferencia entre el pretest —primera medición— y el postest —segunda medición.

Dado que hemos utilizado la misma muestra en las dos mediciones, las características iniciales o experiencia previa de los entrevistados no constituyen rival explicativo. La diferencia que encontremos en el voto no puede ser explicada por características de los entrevistados (sexo, edad, renta, etc....) que a la vez se relacionen con la campaña y con el cambio de voto. Las características de los entrevistados son las mismas antes y después de la campaña, por lo que quedan controladas.

Es evidente, sin embargo, que no están controladas las posibles explicaciones alternativas derivadas de la segunda gran categoría: los acontecimientos externos. Entre la primera y la segunda medición además de la campaña se producen otros sucesos que no controlamos y que además puede que ni siquiera podamos nombrarlos. Puede que durante el periodo de campaña estalle un escándalo político que sea realmente el que produzca el cambio o que tenga lugar algún acontecimiento diferente. Podríamos pensar además en sucesos que afectan de un modo diferente a distintos grupos de nuestra muestra y que pudieran haber producido el cambio puesto de relieve por la diferencia entre las dos mediciones efectuadas.

El propio proceso de cambio interno (maduración) de los individuos que forman parte de la muestra es también una alternativa explicativa. Este proceso madurativo puede haberse estado incubando durante un cierto tiempo y manifestarse en nuestra investigación de modo que las diferencias entre mediciones se deban a este proceso y no a la campaña política. Esta explicación tendrá más plausibilidad si el tiempo transcurrido entre ambas mediciones es suficiente para que se produzca este proceso de maduración o desarrollo interno.

Por último, el cambio puede estar producido por las propias técnicas de medición o recogida de datos utilizadas. En este ejemplo cabe pensar en un posible proceso de aprendizaje que se produce al aplicar el mismo instrumen-

to de medición al mismo grupo de personas en dos momentos distintos; asimismo, el propio instrumento de medición puede haber provocado un cambio en lo que se está midiendo.

El diseño que está sirviendo de ejemplo tiene escasa validez interna, lo que trae como consecuencia que los resultados que obtengamos —cambios entre las mediciones o ausencia de cambios— sean difícilmente interpretables. No sabremos a qué se deben estos cambios o ausencia de cambios al haber distintas explicaciones plausibles y no estar éstas controladas.

¿Cómo lograr un mayor grado de validez interna? La respuesta es aparentemente sencilla: controlando todas las hipótesis alternativas excepto aquella que nos interese a nosotros por constituir el objetivo de la investigación. Este control puede efectuarse *a priori*, es decir, al plantear el diseño de investigación, o *a posteriori* en el análisis e interpretación de los resultados. La utilización de este último método requiere la identificación de las hipótesis alternativas y su introducción en el propio diseño de investigación.

El control *a priori* utiliza esencialmente dos técnicas:

- mediciones repetidas de las variables explicativas y/o controladas que incluyan como mínimo una medición antes y después, y
- grupos de control equivalentes.

El criterio de validez externa

Se dice que un diseño tiene validez externa cuando los resultados obtenidos pueden ser generalizados a otras unidades de análisis, otras situaciones/contextos y otros momentos distintos a aquél en que se realizó la investigación.

La generalización de resultados no se circunscribe a problemas relacionados con muestreo representativo de las unidades de análisis utilizadas; toda investigación pretende, asimismo, generalizar resultados obtenidos en un contexto/situación determinados a diferentes contextos/situaciones de la misma manera que se presente generalizar a otras fechas/momentos distintos de aquellos en que se ha realizado la investigación.

Desde este punto de vista, la perspectiva radical adoptada inicialmente por Campbell y Stanley sobre la imposibilidad de justificar lógicamente la generalización (léase la inducción) debe ser abandonada. Ciertamente Hume mostró hace ya tiempo que la inducción no se podía justificar utilizando la lógica deductiva, pero, claro está, quizás esto no sea necesario y baste con justificarla inductiva y pragmáticamente, admitiendo la justificación que da el sentido común a las numerosas generalizaciones que hacemos cotidianamente.

Para generalizar tenemos que presuponer que conocemos las leyes que lo permiten y desde ese punto de vista sí que conocemos lo relativo a las relaciones entre muestras y poblaciones. Precisamente esto hace que esta dimensión de la validez externa sea más fácil de abordar y resolver.

3. Diseños de investigación social: criterios operativos

Conviene señalar que es la existencia de *efectos interactivos* —o la posibilidad de dicha existencia— lo que hace que los diseños de investigación tengan mayor o menor validez externa. Si nosotros llevamos a cabo una investigación sólo con mujeres donde uno de los resultados fuera que a mayor edad mayores ingresos, podremos generalizar este resultado a los hombres sólo si creemos que no hay efectos interactivos entre sexo, edad y renta o, dicho de otro modo, si la relación edad/renta no es diferente para hombres y mujeres. De la misma manera, si realizamos una investigación en Las Alpujarras sólo podemos generalizar los resultados si nos aseguramos de que no haya variables situacionales o contextuales de dicha comarca que interactúen con nuestros resultados haciéndolos específicos respecto de algún conjunto de condiciones que normalmente están escasamente explicitadas.

La evaluación del grado de validez externa de un diseño de investigación —o de una investigación— es un proceso inductivo de búsqueda de posibles efectos interactivos. *A priori* podemos adoptar cuatro posibles estrategias (Cook y Campbell, 1977):

- a) Utilizar métodos de muestreo aleatorio, tanto en lo que se refiere a unidades de análisis como a contextos/situaciones y tiempo. Obviamente, resulta más sencillo aplicar esta estrategia en la selección de las unidades de análisis que en la de contextos/situaciones; en este último caso sería necesario disponer de una taxonomía o listado de las mismas para poder hacerlo. Esta es, en cualquier caso, la estrategia óptima que minimizaría los problemas de generalización; sin embargo, no siempre es posible recurrir a ella y entonces entran en juego las siguientes posibilidades.
- b) Creación y selección deliberada de *grupos heterogéneos* de unidades de análisis, situaciones/contextos y fechas/momentos. La investigación se llevaría a cabo en contextos muy distintos, con unidades de análisis variopintas y en fechas diferentes; de este modo es posible analizar si los resultados obtenidos se mantienen en las diferentes situaciones, unidades de análisis y momentos, lo que justificaría una eventual generalización de las mismas, avalada por una cuasi representatividad por parte de los diferentes grupos. Esta es la estrategia que se sigue en la selección de muestras para pretestar un borrador de cuestionario. Se busca una muestra en la que estén representados todos los grupos dispares y heterogéneos de una población determinada de modo que podamos contar con sus respuestas al elaborar el cuestionario definitivo. Esta representatividad no suele ser una representatividad conseguida a través de muestras aleatorias, sino a través de muestreos por cuotas forzando la inclusión en la muestra de las unidades de análisis más heterogéneas y distintas.
- c) Utilización de ejemplos/casos *modales*. Si los objetivos de la investigación y los propios conocimientos e información del investigador

sobre el tema a investigar lo permiten, establecer las características de las unidades de análisis, contextos/situaciones y fechas, de modo que sean *modales* para los objetivos o área de investigación. En este caso, la generalización estará más justificada.

- d) Como caso especial de la anterior estrategia cabe pensar simplemente en determinar las características de aquellas unidades de análisis, contextos y momentos a los que se quiere generalizar. Estas características no tienen por qué ser modales, sino simplemente pertenecer o definir aquellas unidades de análisis, etc.... a las que queremos generalizar luego nuestros resultados.

Estas cuatro estrategias están ordenadas de mayor a menor poder de inferencia y también ordenadas en cuanto a posibilidad práctica de utilización; ahora bien, el orden de esta última dimensión es a la inversa que la anterior, de modo que la última estrategia mencionada es la más fácil y práctica de llevar a cabo, pero también la que tiene menor capacidad de generalización, o sea, menor grado de validez externa.

El criterio de validez de constructo

Constituye este criterio un caso especial del criterio de validez externa. Todo investigador opera siempre en dos niveles diferentes:

- un nivel teórico, abstracto, poblado de conceptos y constructos
- otro nivel de observables, de operaciones y de indicadores.

Los dos niveles están estrechamente interrelacionados, puesto que se pretende que uno —nivel operacional— sea traducción del otro —nivel teórico. Pero esta correspondencia o relación epistémica es imperfecta siempre; siempre se operacionaliza con unos grados de libertad. Las operacionalizaciones de constructos y conceptos no son únicas; de aquí que convenga interrogarse sobre el efecto que pudiera tener una diferente operacionalización en los resultados de la investigación.

Traducir el concepto de «clase social» en sentimiento de pertenencia a una determinada clase social a través de una o varias preguntas de un cuestionario, con toda seguridad afectará los resultados de una investigación de modo que éstos serían distintos si la traducción de dicho concepto se hiciera en términos de dimensiones de estatus/educación, ocupación, renta, etc., o si se hiciera en términos de posesión o no posesión de los medios de producción.

En términos de Cook y Campbell, el criterio de validez de constructo tendrá que ver con «la posibilidad de que la definición operativa de una causa o de un efecto pueda construirse en términos de más de un constructo» (Cook y

3. Diseños de investigación social: criterios operativos

Campbell, 1977: 238). Lo que un investigador puede tomar como relación entre clase social y conducta electoral, otro lo podría tomar como relación entre status o prestigio social y conducta electoral. En definitiva, la pregunta que el investigador debe responder es la siguiente: ¿A qué constructos puede generalizarse la operacionalización que nosotros hemos llevado a cabo o tenemos previsto realizar?; o alternativamente: ¿realmente nuestra operacionalización responde a nuestra conceptualización?

El problema de la validez de constructo es claramente una cuestión de medición y de una manera global la solución es utilizar la operacionalización múltiple —múltiples indicadores/múltiples métodos.

Los factores que van en contra de la validez de constructo tienen que ver *bien* con no haber podido incorporar a la operacionalización efectuada todas las dimensiones del constructo investigado (es decir, infrarrepresentación del constructo), *bien* con la utilización de dimensiones irrelevantes en la operacionalización (sobrerrepresentación).

La manera de lograr un alto grado de validez de constructo es:

1. Delimitación clara y precisa de los constructos a utilizar, de modo que salte a la vista de una manera inmediata si la operacionalización resultante responde o no a todos y cada uno de los matices de dicha definición.
2. Utilización de operaciones múltiples en la traducción del constructo.
3. Utilización, siempre que sea posible, de métodos múltiples en la recogida de datos.

El criterio de validez estadística

El análisis de los datos presupone en la mayoría de las investigaciones un análisis estadístico, sea éste descriptivo o inferencial. En este último caso, el investigador tiene que tomar decisiones sobre posibles efectos o relaciones. Debe decidir si existe un efecto o no y, en muchos casos, estimar la magnitud de dicho efecto. En una palabra, al trabajar con muestras, tienen que recurrir a tests de significación. Estos no permiten tomar decisiones de una manera inequívoca, entre otras cosas porque el nivel de significación (probabilidad de acertar o equivocarse) no es algo que esté ya dado y el investigador tiene que decidir a qué nivel trabajará. El «sacrosanto» 5 por 100 ($p = 0,05$) es, a este respecto, una simple costumbre.

Por ello, al utilizar los tests de significación podemos equivocarnos al decidir que no hay un efecto/relación de la misma manera que podemos hacerlo al decidir que hay un efecto/relación. Es por esto por lo que Campbell y Cook consideran este tipo de validez un caso especial de la validez interna utilizando la analogía de sesgo/error aleatorio. La validez interna tendrá que ver con los sesgos, es decir, con los errores sistemáticos, mientras que la validez esta-

dística tendrá que ver, ante todo, con el error aleatorio, es decir, con la fiabilidad y estabilidad de nuestros datos.

Ambos autores señalan, asimismo, dos aspectos relacionados con este tipo de validez que la contextualizan:

1. Es imposible desde un punto de vista lógico probar, o sea, decidir sin error, que no existe determinado efecto o, de una manera más general, la hipótesis nula.
2. La posibilidad de concluir equivocadamente que sí existe covariación es la contrapartida de la posibilidad de concluir equivocadamente que *no* existe covariación entre variables.

La lista de factores que potencialmente pueden afectar la validez estadística de un diseño es bastante grande:

- el poder estadístico del test correspondiente
- la fiabilidad de las medidas
- factores aleatorios del proceso de investigación
- heterogeneidad aleatoria de los sujetos a investigar, etc.

Cómo utilizar los diferentes criterios

Los cuatro criterios de validez señalados se resumen en dos criterios principales (interna, externa), con dos casos especiales cada uno (estadística y de construcción). El intento de aumentar el grado en que un diseño posea un tipo de validez normalmente suele llevar a que disminuya el grado de posesión de otro tipo de validez. Así, el control de explicaciones alternativas (validez interna), normalmente lleva a trabajar con muestras cautivas en experimentación de laboratorio, lo que resulta en un diseño con escasa capacidad de generalización (validez externa) aunque no tiene efectos negativos sobre la validez estadística.

La estrategia a seguir es una en que se logre un compromiso entre los diferentes tipos de validez, de modo que el diseño sea óptimo. En este sentido Campbell y sus colaboradores, Stanley y Cook, insisten una y otra vez en la primacía de la *validez interna*. Si no podemos interpretar unos resultados por no tener controladas las explicaciones alternativas no tiene sentido hablar de errores aleatorios (validez estadística). Por otra parte, si el diseño tiene validez interna, aunque no podamos generalizar, será un diseño que nos dará unos resultados «reales», aunque circunscritos a un entorno restringido. Este es el argumento empleado por Campbell para justificar la primacía de la validez interna.

Ahora bien, conviene destacar dos hechos incuestionables:

1. El tipo de validez que deba primar dependerá del tipo de investigación y de sus objetivos específicos. Una investigación que pretenda prede-

3. Diseños de investigación social: criterios operativos

cir la conducta electoral no tiene por qué preocuparse por la validez interna, sino por la validez de constructo, externa y, en menor medida, por la validez estadística. Del mismo modo, si pretendemos medir el efecto que tiene un programa de ayuda a escolares de una determinada ciudad, la validez externa tendrá escasa importancia, siendo más relevante la validez interna.

Por tanto, la utilización de un tipo de criterio de validez u otro dependerá básicamente de los objetivos de la investigación y habrá que analizarlos cuidadosa y precisamente antes de establecer un diseño de investigación.

2. En muchos casos es posible lograr un adecuado diseño de investigación con alto grado de validez de los cuatro tipos señalados. Por ejemplo, ciertamente la experimentación es el tipo de investigación que permite *a priori* lograr una mayor validez interna. Sin embargo, los diseños cuasiexperimentales² planteados teniendo en cuenta la lógica experimental consiguen generalmente un alto grado de validez interna y tienen, por supuesto, mayor grado de validez externa que muchos experimentos de laboratorio. El desarrollo de los llamados experimentos de campo en psicología social frente a los experimentos clásicos va en esta dirección.

Algunos diseños de investigación

La mayoría de las investigaciones que tienen lugar en sociología son de tipo descriptivo, tratan de dar a conocer cómo es la realidad social de un modo detallado, aproximándose a veces al establecimiento de relaciones entre fenómenos y características de esta realidad. Este hecho viene justificado tanto porque la descripción es normalmente un paso previo para la explicación, como porque normalmente existe un gran desconocimiento de la misma.

Este tipo de investigaciones deben orientarse hacia el logro de *validez externa*, si se trabaja con muestras como suele ser el caso, y la *validez de constructo*, puesto que se trata de explicitar detalladamente la realidad social. Como no hay nada que explicar, no tiene sentido controlar explicaciones alternativas (validez interna) y, del mismo modo la validez estadística queda relegada a un plano secundario.

Más problemático es el planteamiento de diseños de investigación tipo explicativo, puesto que el logro de un diseño adecuado debe responder a los cuatro criterios señalados. Desde el punto de vista de la validez interna se puede trazar un continuum que va desde diseños experimentales a diseños correlacionales, pasando por experimentos de campo, experimentos naturales, diseños cuasiexperimentales y correlacionares de laboratorio. En esta misma línea se puede hablar de diseños preexperimentales, cuasiexperimentales y experimentales.

Resulta mucho más fructífero, sin embargo, describir y analizar ventajas e inconvenientes de diferentes diseños utilizados comúnmente. Veamos algunos de ellos.

1. El «estudio de un caso»

Consiste simplemente en la medición en una sola vez de las características de un grupo o una muestra de unidades de análisis. En el caso de que utilizáramos la población total también se trataría de este mismo tipo de diseño. Es por tanto una medición transversal realizada en una sola vez; obviamente, esta medición puede ser muy compleja, o sea, realizarse un conjunto de mediciones, o incluso en el caso de estudios antropológicos recoger múltiples datos sobre una comunidad. Lo que define este tipo de diseño, denominado por Campbell «diseño postest sin grupo de control equivalente», es precisamente que sólo hay una recogida de datos y que no existe un grupo de comparación, sea éste equivalente o no.

La labor realizada por Lombroso de reunir mediciones de criminales para apoyar empíricamente su teoría del criminal nato constituiría un ejemplo de este diseño, aunque Lombroso realizó la recogida de datos a lo largo del tiempo. De la misma manera, la típica encuesta hecha sobre una determinada población en un momento concreto también sería un diseño postest sin grupo de control.

Si se supone que se está tratando de observar los resultados/consecuencias de un fenómeno concreto anterior a la recogida de datos o medición, el diseño puede esquematizarse, siguiendo a Campbell, como sigue:

$$X \quad 0$$

donde X es el fenómeno/variable/tratamiento cuyas consecuencias queremos analizar y 0 es la medición o recogida de datos.

En el caso de una encuesta o en el caso de los trabajos de Lombroso normalmente no existe un fenómeno cuyos efectos queremos analizar y el esquema sería simplemente

$$0$$

Naturalmente, el simplismo de este diseño muestra ya su gran debilidad en lo que respecta al criterio de validez interna. Sin embargo, conviene recordar algunos puntos importantes:

1. El estudio de casos tipo antropológico lleva emparejado un control importante de explicaciones alternativas que se realiza a través de sucesivas mediciones y de la utilización del método hipotético deductivo en sucesivas aproximaciones.

3. Diseños de investigación social: criterios operativos

2. Aunque la forma de diagramar la encuesta muestra que no existe grupo de comparación, lo cierto es que el investigador tiene siempre una base de comparación que puede ser sobreentendida o real.
3. Normalmente el análisis que se efectúa en el caso de mediciones únicas —casos únicos o encuestas— tiene en cuenta la debilidad inherente del diseño recurriendo a técnicas de control *a posteriori*, mediante modelos multivariantes.

Sin embargo, la debilidad del diseño no queda resuelta por el uso de control estadístico en el análisis ni de grupos reales o aparentes de control. Ciertamente, estos diseños «correlacionales» conllevan usualmente técnicas de análisis cada vez más mixtificadas: análisis de «pautas», ecuaciones estructurales y modelos econométricos, pero este tipo de análisis no resuelve el problema del control de las explicaciones alternativas.

La interpretación de resultados en este tipo de diseños sigue en el mejor de los casos el siguiente proceso:

1. Selección de un modelo empírico que interrelacione las variables explicativas y controladas utilizadas en la investigación congruente con una determinada teoría.
2. Análisis estadístico —*path analysis*, estructuras de covariación, etc.— de los resultados obtenidos de acuerdo con el modelo seleccionado.
3. Decisión sobre si el análisis es congruente con el modelo propugnado o no.

El problema reside en que los datos suelen ser congruentes con más de un modelo, lo que hace que la contrastación empírica sea precaria.

A veces no es posible establecer el orden causal entre variables apriorísticamente, es decir, desde un punto de vista lógico-teórico, no existiendo otro modo de hacerlo, puesto que por definición la medición efectuada es única y simultánea para todas las variables. Además la covariación entre éstas suele constituir un problema en la estimación de los diferentes efectos (colinearidad) y resulta imposible resolver el problema de la espureidad en las relaciones.

Así, no resulta extraño la polémica originada por los estudios de Erlich sobre los efectos disuasores de la pena de muerte en los que utiliza modelos econométricos, es decir, ecuaciones estructurales. Esta cita textual de Friedman resume dicha polémica y también los problemas que, este tipo de diseños correlacionales tan generalizados en sociología, plantean.

El análisis de regresión múltiple no puede nunca proporcionar prueba de que un factor causa otro, aunque puede dar evidencia circunstancial razonable sobre dicha causalidad si: 1) se documenta suficientemente una relación estadística, y 2) todas las explicaciones teóricas plausibles de dicha relación se muestran consistentes, tanto en lo que respecta a la dirección de la causa como del efecto entre los dos factores (L. S. Friedman, 1979: 63).

2. El diseño de un solo grupo con pretest y posttest

Este diseño, diagramado de la siguiente manera:

O_1 X O_2

es una extensión lógica del diseño anterior cuando se trata de medir el impacto de alguna intervención social, programa o cambio. Así sucede, por ejemplo, con las campañas políticas en época de elecciones, cuya evaluación suele hacerse realizando una medición antes y otra después de la campaña o cuando se pretende evaluar el impacto de un programa de televisión o de un cambio en una empresa o institución. Las aplicaciones de este diseño son múltiples y el análisis del mismo giran siempre en torno a las diferencias entre la primera (O_1) y la segunda medición (O_2). Éstas, de existir, se achacan a esa modificación, intervención o programa (X).

La no existencia de un grupo de control, es decir, un grupo que sirva de base de comparación al no haber sido influido por X hace que este diseño, al igual que el anterior, sea inherentemente débil en el control de las alternativas explicativas. No se controlan en absoluto otros hechos o acontecimientos externos al grupo objeto de estudio que han tenido lugar entre la primera y la segunda medición y que pudieran estar relacionadas a la vez con el cambio, programa o intervención y sus resultados; tampoco se controlan los hechos o acontecimientos internos al grupo que tienen lugar entre ambas mediciones y que al igual que antes pueden afectar a cambios y consecuencias/resultados. Estas dos categorías —la historia y la maduración, en palabras de Campbell— son las fuentes de explicaciones alternativas no controladas que hacen difícil de interpretar los resultados obtenidos al utilizar un diseño de este tipo.

Naturalmente, conociendo que las amenazas a la validez interna provienen ante todo de estas dos clases de fuentes, el investigador puede intentar medir/registrar aquellos factores específicos que estando incluidos en estos dos tipos de fuentes se hayan revelado como posibles explicaciones alternativas. Por ejemplo, al medir el impacto de una campaña política habrá que estar al tanto de otros posibles acontecimientos —campañas de partidos contendientes, sucesos políticos durante la campaña, etc....— para controlarlos *a posteriori* en el análisis como explicaciones alternativas plausibles.

Sin embargo, resulta imposible llevar a cabo esta estrategia con todo tipo de posibles variables perturbadoras. Sólo si el grupo de estudio estuviera físicamente aislado tendríamos justificación para pensar que la historia como fuente de invalidez estaría controlada. Pero quedarían las explicaciones provenientes de la maduración.

Una posibilidad más interesante es la de extender este diseño de modo que se realicen varias mediciones antes y varias mediciones después convirtiéndolo así en un diseño de series temporales.

3. Diseño con grupos de control no equivalente

Es una extensión natural del anterior diseño que intenta, introduciendo un grupo de control, evitar la existencia de interpretaciones alternativas de los resultados obtenidos. Sin embargo, no logra introducir un grupo de control auténtico en el sentido de no ser equivalente al grupo de tratamiento.

Goring, funcionario de prisiones inglés utilizó este tipo de diseño para refutar la teoría del criminal nato de Lombroso. Este último se había limitado a registrar los rasgos anatómicos y físicos de los criminales encarcelados deduciendo que los rasgos comunes a todos ellos eran la causa de la criminalidad. Estos rasgos definirían, según Lombroso, al criminal nato. Lo que hizo Goring fue realizar la medición de estos rasgos en un grupo de criminales y así mismo en un grupo de control no criminal, pero con características socio-demográficas (clase social, educación, empleo, edad, etc.) similares. Pues bien, no existían diferencias de rasgos anatómicos o físicos entre ambos grupos y Goring concluyó que no se podía deducir a la vista de ello que hubiera unos rasgos que definirían al criminal nato. Este tipo de estudio se podría diagramar de la siguiente manera:

X	0
	0

Estrictamente hablando, el diseño de Goring no se ajusta al que se está analizando, que además suele utilizar pretest y postest, pero sí sirve para ilustrar el empleo de grupo de control no equivalente.

Aparentemente, ambos grupos —el criminal y el no criminal— han sido igualados en una serie de variables socio-demográficas, pero esto no les hace equivalentes, sino simplemente más parecidos dado que no se utilizaron más que unas pocas variables para la igualación.

Los resultados obtenidos por Goring echaron por tierra las tesis de Lombroso y esto mismo puede suceder en otros casos en que se utilice este tipo de diseño con pretest y postest. A veces, los propios resultados obtenidos permitirán una fácil interpretación de los mismos.

Por otra parte, en la medida que para la formación/selección del grupo de control se utilicen variables, que razones teóricas o investigaciones anteriores hayan señalado como relevantes, más fácil será la interpretación de resultados de una manera inequívoca. Mediante este mecanismo habremos transformado las variables perturbadoras conocidas en variables controladas y resultará más difícil que haya otras variables perturbadoras no controladas.

4. El diseño de «discontinuidad en la regresión»

En los últimos diez años este tipo de diseño ha experimentado un gran auge sobre todo en el campo de la investigación evaluativa. Han aumentado tanto

las aplicaciones empíricas del diseño como los métodos de planeamiento del mismo y técnicas de análisis. Aunque su campo de aplicación no es muy amplio, sin embargo, tiene la virtud de ser un diseño casi experimental que permite en una aplicación óptima un alto grado de validez interna.

Este tipo de diseño supone:

- una medición pretest
- una medición posttest
- la aplicación de una intervención o un programa, y
- una selección de grupos en base a una característica determinada.

De un modo esquemático, puede diagramarse como sigue:

0	X	0
0		0

Sin embargo, este diagrama puede conducir a error si no se tiene en cuenta cómo son seleccionados los grupos, que es precisamente lo que le diferencia del diseño anterior. Los grupos son seleccionados en función de alguna característica de modo que aquellas personas que la tienen —o que no la tienen— pasan a recibir el programa o intervención, mientras que el resto no lo recibe formando así el grupo de control no equivalente.

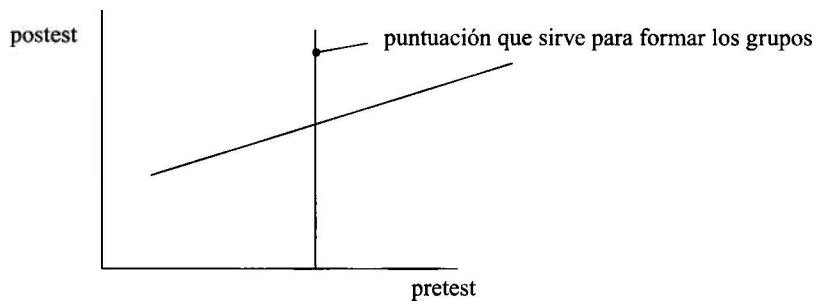
El requisito imprescindible para este tipo de diseño es que exista una característica común a los dos grupos que pueda ser medida a ser posible como variable continua en función de la cual se formen los dos grupos, de modo que todas aquellas personas que estén por debajo —o por encima— de una puntuación determinada pasen a recibir el programa o intervención y todas las demás no. Por ello, este tipo de diseño se utiliza normalmente cuando se forman grupos en función de méritos o necesidades de la gente, como sucede en la educación compensatoria, en medicina al utilizar nuevas técnicas de cirugía, al asignar becas, etc.

El nombre del diseño —discontinuidad en la regresión— proviene del tipo de análisis que se utiliza en el que se efectúa un análisis de regresión de las medidas posttest en las medidas pretest.

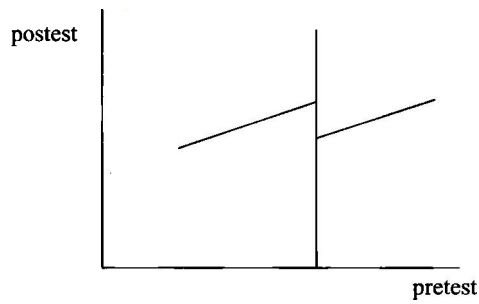
Esta regresión se efectúa por separado para el grupo que recibe el tratamiento/programa y para el grupo de control no equivalente; sólo si existe una discontinuidad en la regresión se podrá hablar de que efectivamente el tratamiento/programa ha tenido un efecto determinado.

Los gráficos adjuntos muestran diferentes posibles resultados de un diseño de este tipo y su interpretación:

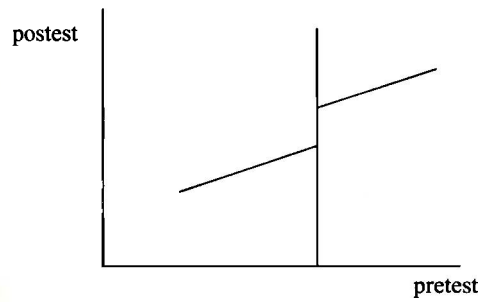
3. Diseños de investigación social: criterios operativos



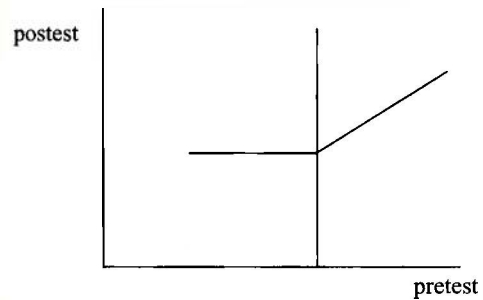
a: No hay efecto



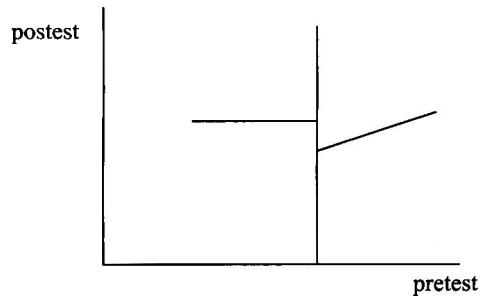
b: Efecto principal positivo



c: Efecto principal negativo



d: Efecto interactivo positivo



e: Efecto principal e interactivo positivos

* Los gráficos están tomados de W. M. K. Trochim, 1984.

Este tipo de diseño será perfectamente aplicable a situaciones en que se quisiera evaluar la rehabilitación o tratamiento de drogadictos, puesto que la asignación del tratamiento se realiza en función de las necesidades y también en el caso de tratamientos psicológicos o psiquiátricos.

5. Series temporales interrumpidas

Como veíamos antes, intentar evaluar una campaña política mediante un diseño de un solo grupo con pretest/posttest resulta imposible, pues entre otras cosas no se controlan las posibles alternativas explicativas relacionadas con la historia, es decir, otros sucesos aparte de la campaña que tengan lugar entre la primera medición y la segunda medición.

Añadir un grupo de control equivalente permitiría evitar estos problemas, pero ¿cómo encontrar un grupo de control cuando la campaña política normalmente afecta a toda la población? Caben dos alternativas:

- a) Utilizar diferentes niveles de exposición a la campaña para formar grupos lo más equivalentes entre sí excepto en dicha exposición de modo que puedan servir de base de comparación mutua, y
- b) extender el número de mediciones de modo que el diseño pretest/posttest se transforme en un diseño de series temporales interrumpidas tal y como aparece en el siguiente diagrama:

$0_1 \quad 0_2 \quad 0_3 \quad X \quad 0_4 \quad 0_5 \quad 0_6$

Ciertamente, desde un punto de vista lógico la historia no queda controlada mediante esta extensión de las mediciones pretest y posttest, puesto que, además de X puede ocurrir alguna otra cosa que realmente sea la causante de la diferencia entre mediciones posttests y pretests. Pero al disponer de varias

3. Diseños de investigación social: criterios operativos

mediciones, un análisis de la tendencia antes y después del suceso, cambio o intervención (X) permite la mayoría de las veces controlar la historia *a posteriori*, pero no obviamente en el propio diseño. Precisamente, esta es la razón por la que se utilizan más de dos medidas pretest/postest.

Uno de los primeros ejemplos de utilización de este tipo de diseño tuvo lugar al principio de los sesenta en la discusión sobre las consecuencias de la pena de muerte. Utilizando este diseño las tesis abolicionistas mostraron que la abolición de la pena de muerte no llevaba aparejado un incremento del número de homicidios, argumento utilizado por los que se oponían a dicha abolición. Este diseño se utilizó con grupos de control no equivalentes comparando las tasas de homicidio de diferentes estados de los Estados Unidos antes y después de la abolición, así como estados con y sin pena de muerte.

Este tipo de diseño fue formalmente «inventado» por Campbell y Stanley e inicialmente parecía tener escasa validez interna. Sin embargo, ésta puede aumentarse no sólo añadiendo un grupo de control, sino también:

1. Recogiendo datos de la serie temporal en intervalos de tiempo muy cortos de modo que sea prácticamente imposible hipotetizar la existencia potencial de causas alternativas.
2. Midiendo o recogiendo información de otras posibles causas alternativas para utilizarlas en un control *a posteriori* de tipo estadístico.
3. Utilizando ajustes cíclicos y estacionales previos en el análisis de las series temporales para evitar precisamente este tipo de explicaciones alternativas.

El trabajo de Ross (1982) es una buena muestra de este tipo de estrategias. Ross utiliza datos de diferentes países sobre accidentes de tráfico antes y después de penalizar más o menos severamente el conducir habiendo bebido para obtener medidas del impacto de esta penalización sobre la conducta. En vez de utilizar las series temporales originales hace uso sucesivamente de:

- a) Correcciones cíclicas y estacionales.
- b) Datos de accidentes de tráfico en aquellos días en que es más normal que existan conductores que hayan bebido (viernes o sábados por la noche).
- c) Distinguiendo entre accidentes mortales y no mortales, por el día y por la noche.
- d) Utilizando ajustes no lineales en el análisis de los datos.

Todo ello permite llegar a conclusiones internamente válidas.

A este respecto conviene señalar que el análisis de los resultados de este tipo de diseños no debe realizarse utilizando mínimos cuadrados normales, sino utilizando el modelo ARIMA, es decir, modelos autorregresivos e integrados de medias móviles. Esta serie de modelos tiene en cuenta el principal

problema de las series temporales, que es la presencia de autocorrelación en los datos.

6. Diseños experimentales

Un diseño experimental auténtico se define por dos características interdependientes:

- la utilización de manipulación/tratamiento de la realidad social por parte del investigador,
- el control de explicaciones alternativas a través de la utilización de al menos un grupo de control equivalente al grupo experimental.

El control se consigue recurriendo a la formación aleatoria de los grupos (experimental y de control) sin que esto implique la no utilización previa de igualación de grupos en aquellas variables identificadas *a priori* como variables perturbadoras. Es decir, en un experimento bien diseñado y ejecutado, las variables perturbadoras más relevantes quedan controladas mediante la igualación de los grupos experimental y de control y el resto de las variables no explicativas quedan aleatorizadas mediante la formación aleatoria de los grupos, es decir, mediante la asignación aleatoria de los sujetos a los diferentes grupos.

Naturalmente, lo que permite llevar a cabo este control es la existencia de la manipulación o intervención del investigador. Éste proyecta una intervención modificadora de la realidad y trata de medir y analizar las consecuencias/efectos de dicha intervención. La planificación de la intervención permite la formación de grupos aleatoriamente al aplicarla a quien se quiera, donde se quiera, como se quiera y cuando se quiera. Por esto, las dos características que definen la experimentación ideal son interdependientes empíricamente, siendo una condición indispensable de la otra (sin manipulación no hay control potencial y al revés).

La lógica experimental resulta perfectamente evidente partiendo de estos dos requisitos ya definidos:

1. Primero se forman al menos dos grupos que sean iguales en todas sus características recurriendo a la asignación aleatoria de sujetos a ambos grupos.
2. Uno de los grupos «sufre» la intervención del investigador mientras que otro no. La determinación de qué grupo será el experimental y cuál el de control se realiza asimismo aleatoriamente.
3. Se comparan los dos grupos una vez realizada la intervención/tratamiento experimental. Puesto que inicialmente los grupos están igualados, cualquier diferencia entre ambos después del tratamiento sólo podrá deberse a dicho tratamiento/ intervención.

3. Diseños de investigación social: criterios operativos

Este es el diseño experimental más sencillo en el que existe un grupo de control y sólo mediciones postests, tal y como muestra esquemáticamente el diagrama

Grupo experimental	Tratamiento	Medición
Grupo de control		Medición

Este diseño puede ampliarse a través de la:

- utilización de más de dos grupos
- realización de mediciones antes del tratamiento.

En el primer caso están comprendidas dos situaciones que se producen con mucha frecuencia:

- a) Cuando se utiliza más de un grupo de control para evitar el efecto placebo. El efecto de un tratamiento específico en determinados casos difícil de aislar del efecto de aplicar cualquier tratamiento; por ello, se recurre a tres grupos:
 - uno experimental *sensu estricto*
 - otro pseudo-experimental con un tratamiento aparente
 - un grupo de control auténtico.

Así, en la experimentación farmacológica puede que la mera administración de un «fármaco», sea cual sea éste, logre una mejoría. Este es precisamente el efecto placebo. Por ello, un grupo recibe el fármaco que se está investigando, otro un fármaco aparente que no tiene efectos reales y otro grupo sirve de estricto control, no recibiendo ninguna medicina.

- b) El tratamiento/intervención puede no ser del tipo o hay tratamiento o no lo hay, de modo que muchas veces hay varios niveles de tratamiento/intervención. En estos casos habrá diferentes grupos experimentales dependiendo del número de niveles de tratamiento. Si el tratamiento viene definido por más de una variable independiente también tendremos varios niveles y, por tanto, varios grupos experimentales.

En ambos casos (varios niveles del tratamiento o más de una variable independiente) no resulta ya estrictamente necesario el grupo de control, puesto que los diferentes grupos experimentales sirven de control/comparación entre sí mismos, aunque los objetivos de la investigación dictarán la necesidad o no de grupo de control en cada caso.

Respecto a las mediciones efectuadas, las mediciones pretest permiten asegurarse de la premisa básica de igualdad de partida de los grupos utiliza-

dos, así como estimar mejor los efectos del tratamiento. Pero contienen un peligro cuya importancia relativa dependerá de cada investigación en concreto. Pueden introducir variables perturbadoras a través de una posible reactancia en la medición y a través del efecto de aprendizaje. Dentro de los diseños experimentales clásicos el diseño de pretest/postest con grupo de control es un claro ejemplo de este tipo de posibilidades. Este diseño puede diagramarse esquemáticamente como sigue

Medición	Tratamiento	Medición
Medición		Medición

Como puede verse, hay:

- formación aleatoria de los dos grupos
- medición antes del tratamiento, y
- medición después del tratamiento.

La lógica experimental permite otros muchos tipos de diseños (Salomón, factoriales, de cuadrado latino o greco-latino, etc....) que presentan la misma línea argumental propia de la experimentación tal como ha sido presentada aquí. El lector interesado puede ampliar este apartado en otros manuales más especializados.

A modo de epílogo

Los diseños analizados en este capítulo forman parte de una amplia trilogía de diseños *correlacionares*, *experimentales* y *cuasiexperimentales*. En sociología se utilizan sobre todo diseños correlacionares que tal y como se ha visto son intrínsecamente vulnerables al no ser posible en la mayoría de los casos una interpretación unívoca de sus resultados. Sin embargo, muchas veces podrían llevarse a cabo diseños cuasiexperimentales con escaso esfuerzo adicional consiguiendo claramente resultados mucho menos equívocos. Dado el predominio de los diseños correlacionales no resulta extraño la falta de acumulación de resultados y la mezcla de ideología y voluntarismo existente en la disciplina.

Ciertamente, existen investigaciones en las que resulta muy difícil diseñar *a priori* ni experimentos ni tan siquiera cuasiexperimentos. Pero un investigador atento puede añadir aspectos a dichas investigaciones que sirvan para controlar todas las posibles variables perturbadoras o al menos gran parte de ellas. Este es el sentido, y no otro, de esta incursión por la lógica de la validez interna, externa, de constructo y estadística. En el planeamiento de una investigación, el control de hipótesis alternativas puede llevar a diseños «parcheados», a «retazos» que formalmente son poco elegantes, pero que son el resul-

3. Diseños de investigación social: criterios operativos

tado de un compromiso entre las posibilidades del investigador y los criterios de validez.

El estudio sobre el efecto del paso por una Escuela Militar mencionado por Campbell y Stanley (1967), sigue constituyendo un buen ejemplo de este tipo de aproximación de diseño «parcheado».

En él los autores fueron añadiendo grupos de control no equivalentes, mediciones pretest/postest y comparaciones con grupos no sometidos a los cursos de la Escuela Militar para lograr el control de la mayoría de las posibles explicaciones alternativas. Estas transformaciones del diseño inicial se efectuaron *a posteriori* y por eso se trata de auténticos «parches», pero el resultado fue un diseño cuasiexperimental con un amplio grado de validez interna.

Esta «lógica del control» debe aplicarse de un modo flexible e imaginativo de modo que las soluciones no pueden en muchos casos formalizarse *a priori*, aunque sí se pueden deducir «técnicas» generales a utilizar como mediciones pretest/postest, comparación con grupos de control equivalentes, igualación de grupos, grupos de control formados en base a diferentes niveles de la(s) variable(s) independientes(s), etc.... Flexibilidad e imaginación en la utilización de los criterios de validez desarrollados en este capítulo es la primera y más importante conclusión que debe extraerse.

Por otra parte, la estructura cognitiva utilizada aquí está fuertemente fundamentada en el artículo seminal de Campbell y Stanley y trabajos posteriores en la misma dirección. Sin embargo, esta estructura cognitiva no es fácilmente aplicable a determinados diseños de investigación muy utilizados en ciencias sociales. Tal es el caso de la observación participante o del estudio en profundidad y durante cierto tiempo de un solo caso, sea éste una institución, grupo, etc...

¿Cómo conceptualizar este tipo de estudios? Podría pensarse que se trata de la medición única aunque muy compleja de un solo caso/muestra o grupo y entonces sería simplemente en la notación de Campbell similar por tanto a la encuesta en la que también se produce una sola medición.

Ahora bien, en la observación participante ciertamente no hay en principio un tratamiento/manipulación pero sí repetidas observaciones a lo largo del tiempo en un contexto natural con múltiples posibilidades de comparación entre grupos más o menos equivalentes y con una potencialidad de falsamiento de hipótesis y teorías muy grande.

A la vista de esto quizá pudiera diagramarse este tipo de investigación de la manera siguiente:

OX OX OX OX OX

respondiendo así a ciertos diseños en que se mide repetidamente una consecuencia de algún suceso/tratamiento, pero tampoco este esquema reflejaría exactamente este tipo de investigación.

El investigador que lleva a cabo observación participante elabora paulatinamente explicaciones de los fenómenos observados a través de la experiencia personal y de la contrastación cotidiana de esas explicaciones con la realidad de modo que al final del proceso investigador los modelos explicativos alcanzados reflejan supuestamente la realidad estudiada. Ahora bien, la escasa formalización del proceso por el que el investigador llega al resultado que llega es la gran debilidad de este diseño en el que otro investigador pudiera llegar a otro tipo de resultados diferentes sin que probablemente pudiera elegirse entre dos interpretaciones alternativas de un modo racional y objetivo.

Notas

1. Dejo fuera de este capítulo el análisis de los diseños exploratorios cuyo objetivo es la formulación más adecuada de un diseño de investigación.
2. Más adelante se define qué es un experimento; un cuasi-experimento significa a la vez un intento de seguir a rajatabla la lógica experimental y una falta de alguna de las características del experimento.

Bibliografía

- Alvira Martín, F. (1984): «La investigación sociológica», en Salustiano del Campo (ed.), *Tratado de Sociología*, Madrid, Taurus.
- (1985): «La investigación evaluativa: una perspectiva experimentalista», *REIS*, núm. 29, Madrid.
- (1985): «La investigación social y la planificación de Servicios Sociales», en *Planificación de Servicios Sociales*, Madrid, Colegio de Ciencias Políticas y Sociología.
- Alvira Martín, F., Avia, M. D., Calvo, R., y Morales, F. (1980): *Los dos métodos de las ciencias sociales*, Madrid, CIS.
- Campbell, D. T. (1984): «Can an Open Society be an Experimenting Society», ponencia presentada en el encuentro sobre Filosofía de Karl Popper, 6 a 9 de noviembre, Madrid.
- Campbell, D. T., y Stanley, J. (1967): *Diseños experimentales y cuasiexperimentales en investigación educativa*, Buenos Aires, Amorrortu.
- Cook, T. D., y Campbell, D. T. (1977): «The design and conduct of quasi-experiments and true experiments in field setting», en H. M. Dunette, *Handbook of Industrial Psychology*, Chicago, Rand McNally.
- Friedman, L. S. (1979): «The use of multiple regression analysis to test for a deterrence effect of capital punishment», en S. L. Messinger y E. Bittner (eds.), *Criminology Review Yearbook*, 1, Thousand Oaks, Calif., Sage.
- Kish, L. (1980): «Representación, aleatorización y control», en F. Alvira y otros, *Los dos métodos de las ciencias sociales*, Madrid, CIS.

3. Diseños de investigación social: criterios operativos

- McTavish, D. G., y otros (1977): «Assessing research methodology», *Sociological Methods and Research*, 6 (1).
- McDowall, D. C., y otros (1980): *Interrupted time series analysis*, Thousand Oaks, Calif., Sage.
- Reichardt, C. S., y Cook, T. D. (1979): *Qualitative and Quantitative Methods in Evaluation Research*, Thousand Oaks, Calif., Sage.
- Ross, H. L. (1982): «Interrupted time series studies of deterrence of drinking and driving», en J. Hagan, *Deterrence reconsidered*, Thousand Oaks, Calif., Sage.
- Selltiz, C., y otros (1981): *Métodos de investigación de las relaciones sociales*, Madrid, Rialp.
- Spector, P. E. (1982): *Research Designs*, Thousand Oaks, Calif., Sage.
- Trochim, W. M. K. (1984): *The Regression-Discontinuity Approach*, Thousand Oaks, Calif., Sage.
- Tromkin, H. L., y Strenfert, S. (1977): «Laboratory Experimentation», en H. M. Dunnette, *Handbook of Industrial Psychology*, Chicago, Rand McNally.