ANUARIO DE PSICOLOGIA
NUM. 21 - 1979 (2)

PROBLEMATICA METODOLOGICA DE LA VALIDEZ EXPERIMENTAL

JAUME ARNAU GRAS

Departamento de Psicología Experimental
Universidad de Barcelona

Es evidente que uno de los aspectos más importantes que tiene planteada la investigación experimental consiste en la «obtención de inferencias válidas sobre la relación de variables, basadas en cuidados/as observaciones de la naturaleza» (Rosenthal y Rosnow, 1975, p. 93). Guiados, precisamente, por este propósito los investigadores han llegado a la conclusión que se pueden obtener «inferencias válidas» en un mayor porcentaje cuando se han aplicado controles adecuados que cuando éstos no han estado presentes. No es menos cierto, sin embargo, que se han ejecutado experimentos, altamente controlados, perfectamente diseñados, pero cuyos resultados no han podido ser aplicados a ningún tipo de universo. Y es precisamente este difícil equilibrio entre la exigencia del control y la necesidad de descubrir relaciones constantes y de gran alcance entre las variables, lo que ha llevado a los experimentadores a plantearse el problema relativo a la validez. Temática que ha sido puesta de manifiesto inicialmente por Donald T. Campbell (Cambell y Stanley, 1962; Webb, Campbell, Schwarts y Sechrest, 1966), y que actualmente preocupa a la mayoría de los experimentadores.

1. VALIDEZ DEL EXPERIMENTO

Para tener una idea general de lo que representa el concepto de validez dentro del contexto de la investigación experimental, vamos a utilizar la siguiente representación diagramática de la hipótesis experimental:

variable independiente→relación→variable dependiente
valores de otras variables

Toda hipótesis experimental describe un tipo de «relación causal» entre «dos o más variables». De ahí que toda hipótesis incluya una relación así como los términos o extremos a los que dicha relación se refiere. Un experimento será válido en la medida que nos permita, a partir de sus datos, inferir y mantener una determinada hipótesis para un universo concreto de cosas y objetos. En caso contrario, el experimento pierde todo su sentido y carece de significado: es decir, no sirve para nada. Se puede, por tanto, afirmar que en términos generales un experimento es válido si permite distinguir «entre una hipótesis plausible y otra rival falsa, con un mínimo grado de error». En la medida, pues, en que una supuesta relación se cumpla para un determinado nivel de referencia objetiva, queda definida la validez del experimento a partir del cual se ha obtenido dicha evidencia. De ahí que todo experimento debe

recoger los dos aspectos básicos que implica el concepto de validez: interna y externa. Como señala Gottsdaner (1978, p. 43), la «validez interna» se refiere a la relación en sí misma. Se dice, pues, que un experimento tiene validez interna siempre que la relación enunciada en la hipótesis se infiere del mismo en función del valor de las variables que han sido utilizadas, así como para los niveles con que han actuado las restantes. La validez interna implica, pues, que el experimento se haya planteado correctamente, permitiendo, al mismo tiempo, inferir la relación propuesta en la hipótesis experimental. En cambio, la validez externa se refiere, fundamentalmente, a las variables implicadas en el experimento y por extensión, a los diferentes niveles con que pueden actuar las restantes variables extrañas. En la medida que un experimento recoja la amplia variabilidad de las distintas variables directa o indirectamente relacionadas con el mismo, se puede afirmar que dicho experimento es adecuado. Cuanta mayor amplitud y alcance posea el diseño experimental, más potentes serán los resultados obtenidos del mismo.

De la discusión de la validez interna y externa del experimento, podemos, por tanto, concluir que cuando un experimento ofrece la posibilidad de probar, con total garantía, una hipótesis teórica, dicho experimento posee validez interna. Por otra parte, si dicha relación que ha podido ser comprobada con un alto nivel de precisión y exactitud, no puede verificarse en otros laboratorios, por otros investigadores, con otros sujetos, o en diferentes contextos, la relación descubierta y fielmente comprobada posee escasa o nula vigencia; es una relación de alcance limitado o nulo. Tiene, en definitiva, poca validez externa.

2. VALIDEZ INTERNA Y CONTROL

El carácter básico de la investigación experimental consiste en proporcionar una base segura que permita comprobar la «relación causal» entre la variable independiente y la dependiente. De ello depende, como hemos visto, la validez interna del experimento; es decir, la posibilidad de inferir conclusiones precisas y exactas sobre la relación entre las variables experimentales. Para conseguir dicho propósito es necesario mantener las «restantes variables relevantes independientes» a un nivel de acción constante a lo largo del experimento. Sólo así se puede alcanzar el «experimento ideal» que según Keppel (1973, p. 23) sería aquel en el que sólo es permitido variar la variable independiente. Cualquier otra variable debe ser mantenida constante, de forma que la variable de medida sólo sea afectada, en su variabilidad o cambio, por la acción de los tratamientos de la variable independiente.

Una de las principales dificultades que suelen presentarse en algunas situaciones experimentales es el hecho de que una variable relevante extraña varíe o cambie de valor simultáneamente con la variable manipulada. Ante

una situación así cabe preguntarnos si la causa de los cambios observados en la variable de medida se deben a la manipulación de la variable independiente o si por el contrario, son resultado de la presencia de una variable extraña contaminadora. De ello se desprende la importancia que adquieren este tipo de variables contaminantes, denominadas por Borkowski y Anderson (1977, p. 116) «variables de confundido». Es obvio, por otra parte, que si un investigador ha realizado un experimento en el que ha intervenido una variable de confundido y ésta no ha sido adecuadamente controlada, cualquier conclusión que pretenda derivar deberá ser cuestionada, y, consiguientemente, rechazada por falta de validez interna. No controlar adecuadamente la acción de las variables de confundido es atentar gravemente contra la validez interna del experimento.

Por otra parte, hemos de reconocer que la labor de detectar la presencia de variables de confundido es difícil y ardua, y algunas veces puede pasar desapercibida su presencia, incluso, para investigadores con gran experiencia en trabajos experimentales. Y es tal la dificultad de dar con dichas variables que, históricamente, se han aceptado datos de experimentos contaminados durante un largo período de tiempo, hasta que se ha podido demostrar la presencia de una variable de confundido. Tal es el caso recogido por Borkowski y Anderson (1977), sobre un experimento realizado por Collier y Marx (1959). En dicho experimento, Collier y Marx (1959) se propusieron comprobar el efecto de «una previa exposición a una determinada cantidad de recompensa» sobre la ejecución futura de los sujetos. A tal efecto formaron tres grupos de ratas, con características lo más similares posibles, sobre todo en lo relativo a la edad y peso, para garantizar la homogeneidad de los grupos. El experimento se dividió en dos partes. En la primera parte —fase I del experimento— los tres grupos de ratas recibieron diferentes soluciones de agua azucaradad: 4 %, 11,3 % y 32 % de concentración de azúcar respectivamente. Esta fase duró una serie repetida de sesiones, pasándose a continuación a la segunda fase del experimento (fase II). En esta segunda fase, los animales tuvieron que realizar una respuesta de presión de barra a fin de conseguir una recompensa líquida por minuto. Dicha recompensa, la misma para todos los sujetos, consistía en una solución azucarada con una concentración del 11,3 %. Obviamente esta segunda fase introducía un cambio de recompensa en dos grupos del experimento, el primero y el tercero. De esta forma los grupos con más alta o baja recompensa cambiaban, en el período de entrenamiento de presión de barra, a un nivel de concentración inferior o superior. En cambio, el segundo grupo, se mantenía, a lo largo del experimento, con la misma cantidad de recompensa; una cantidad de 11,3 % de concentración azucarada. Con exclusión de la variable manipulada —concentración de azúcar— Collier y Marx controlaron las restantes condiciones que pudiesen afectar diferencialmente la ejecución de los tres grupos.

Los resultados de dicho experimento demostraron que, sobre un total de 10 sesiones para la segunda fase o fase de entrenamiento de presión de barra

para la obtención de una cantidad de recompensa de 11,3 %, la tasa de presión de barra difería considerablemente en los tres grupos. Así el grupo que cambió a una tasa más alta, grupo de 4 %, presentó una tasa de presión muy alta; en cambio el grupo de 32 % presentó la tasa más baja, en comparación con el grupo que no experimentó cambio alguno o grupo control. Collier y Marx interpretaron dichos resultados en el sentido que el efecto de una determinada cantidad de recompensa puede ser alterado «por la previa experiencia de una cantidad diferente de la misma recompensa». Esta interpretación, basada fundamentalmente en el concepto de contraste, fue aceptada durante unos diez años hasta que Dunham y Kilps en 1969 comprobaron que los datos reunidos por Collier y Marx no se relacionaban con la variable manipulada —cambios en la cantidad de la recompensa— sino que posiblemente se debían a la presencia de una variable de confundido. De esta forma ejecutaron un experimento en el que se verificó que las diferencias en los pesos medios que desarrollaron las ratas durante la primera fase tenían un efecto en la ejecución que presentaban en la segunda fase. En efecto, el grupo con una cantidad de recompensa del 32 % de concentración presentaba un relativo incremento en su peso medio, siendo lo contrario verdadero para el grupo de 4 % de recompensa, cuando eran comparados con los sujetos del grupo control que no experimentaron cambio de peso alguno. De acuerdo con la interpretación de Dunham y Kilps estas diferencias de peso podían haber persistido a lo largo de la fase de entrenamiento de presión de barra, aun cuando todos los grupos hubiesen pasado a un mismo nivel de concentración de azúcar. De ello podía inferirse que las diferencias entre los pesos promedios de las ratas de los grupos experimentales covariaban con las diferentes condiciones de tratamiento, por cuya razón se la podía considerar como una variable de confundido.

Ante estos resultados cabe la posibilidad de preguntarnos si la exposición a las diferentes concentraciones de azúcar durante el período de preentrenamientos, o las diferencias de peso que desarrollaron los grupos durante esta fase, o ambas cosas a la vez, fueron la causa de los resultados obtenidos por Collier y Marx. A tal efecto Dunham y Kilps llevaron a cabo el experimento al que nos hemos referido y en el que se mantuvieron iguales los pesos de los cuerpos. De esta forma se comprobó que las tasas diferenciales de presión de barra en función de los cambios de recompensas no se dieron en la segunda fase del experimento, en contra de los resultados obtenidos por Collier y Marx. De esta forma Dunham y Kilps concluyeron que la variable que Collier y Marx habían explícitamente manipulado, cambios de recompensa entre la primera y segunda fase del experimento, no constituía la causa de las tasas diferenciales de presión de barra producidas durante el entrenamiento. Sí, en cambio, pudo inferirse que la variable confundidora —diferentes pesos de cuerpo desarrollados durante la primera fase— fue la variable funcional.

El experimento de Collier y Marx sirve de ejemplo para ilustrar cómo pueden interpretarse erróneamente unos datos experimentals a causa de la

presencia de una inadvertida variable de confundido. Nos demuestra, también, que detectar una posible variable de confundido, es tarea difícil y muchas veces, como en el caso a que nos hemos referido, permanece oculta durante un largo período de tiempo. A pesar de ello, los resultados de Collier y Marx, erróneamente atribuidos a una variable manipulada, ejercieron un notable influjo en el área de los efectos de refuerzo por contraste.

La lección básica que cabe derivar de este ejemplo histórico es que todo experimento, correctamente diseñado, exige la localización y eliminación de las variables de confundido. Por ello consideramos que un estudio y análisis de las posibles fuentes de confundido que pueden darse en toda situación experimental constituye un ejercicio metodológico útil, ya que nos va a permitir localizar con mayor facilidad los factores de confundido que son más probables así como su correcta identificación.

Diversos autores han propuesto distintas clasificaciones de las principales fuentes de confundido y en base a ellas vamos a proponer su estudio y clasificación (Underwood, 1957; D'Amato, 1970; Meyers y Grossen, 1974; Boskowski y Anderson, 1977; Gottsdanker, 1978).

3. FUENTES DE CONFUNDIDO

La forma cómo una variable puede distorsionar la interpretación de un experimento depende, básicamente, de su origen. Las principales fuentes de confundido proceden de los componentes básicos de que se halla integrada la situación experimental e incluyen las siguientes: (a) Variaciones inadvertidas de algunos aspectos de la situación o contexto en el que se ejecuta el experimento; (b) Variaciones inadvertidas que se producen en los propios sujetos que participan en el mismo, y (c) Variaciones inadvertidas que proceden del propio experimentador. Situación, sujeto e investigador constituyen las principales fuentes de confundido dado que, en mayor o menor grado se hallan presentes en todo diseño experimental.

3.1. Variables de confundido situacionales

Entre las variables situacionales que constituyen en sí mismas el marco de referencias externo del experimento y a las que forzosamente se relaciona cualquier experimento, se pueden señalar las siguientes: (1) Variables ambientales; (2) Variables de tarea y (3) Variables de procedimiento. Cada una de estas tres categorías agrupa una serie de potenciales variables de confundido que, en el caso de que inadvertidamente covarían con la variable manipulada, pueden constituir la causa de los cambios operados en la variable de medida o dependiente.

3.1.1. *Variables de confundido ambientales*

Las variables de confundido ambientales abarcan toda la gama de factores o aspectos que se hallan presentes en el contexto ambiental en el que se inscribe el experimento. Cualquier cambio que se produzca en aspectos tales como iluminación, ruido, temperatura, humedad, etc., puede convertirse en la auténtica causa de la ejecución registrada u observada en los sujetos experimentales. Es evidente que si no se aplican los adecuados controles, estas variables pueden actuar tanto como elementos motivadores o distractores de la tarea a ejecutar. De ello se deriva la necesidad y urgencia de un adecuado control de estas fuentes de variación a fin de neutralizar y en el caso que ello sea posible, eliminar su potencial acción.

Entre las técnicas de control de este tipo de variables de confundido cabe destacar tres que, tradicionalmente, han sido consideradas como procedimientos básicos de control y reciben los nombres siguientes: (a) Mantenimiento constante de la misma; (b) Sistematización y (c) Aleatorización.

Una de las técnicas de control más frecuentes, sobre todo en lo que se refiere a las variables ambientales, es el «mantenimiento constante de las mismas». En el supuesto que dicha técnica haya sido aplicada de forma efectiva, se puede asumir que su influencia se mantiene constante a través de todos los sujetos y grupos experimentales. Por lo general este tipo de control se consigue mediante la construcción de especiales locales de laboratorio o cabinas en las que sea posible un total y efectivo control de las principales fuentes de confundido ambientales. Una forma ejemplificada de control ambiental la tenemos en la famosa «caja de Skinner» que, como sabemos, ha sido ideada a propósito para obtener el más completo control de todas las posibles variables de confundido ambiental.

También la sistematización puede, en algunos casos, constituir una eficaz técnica de control. En base a su sistematización, la variable independiente extraña, objeto de control, se convierte en variable independiente y es tenida en cuenta en el momento de analizar los datos. Hemos de reconocer, a pesar de esto, que una de las formas más efectivas de control para las variables ambientales son las técnicas mecánicas como las cabinas, gabinetes, cajas, etc., mediante las cuales podemos regular perfectamente la temperatura, luminosidad, ruido y las restantes variables de confundido. Ahora bien, no es menos cierto que en algunos casos este tipo de control es totalmente imposible, por cuya razón se hace aconsejable su sistematización. Fijémonos en la variable «hora del día en que se ejecuta el experimento». Esta variable, si se demuestra que tiene una influencia efectiva sobre los resultados, debe ser de alguna forma controlada. Dado que de lo contrario, nos expondríamos a obtener unos resultados confundidos.

No es menos cierto, sin embargo, que la mayoría de las variables ambientales no suelen presentar un gran interés teórico para la mayoría de los experimentos conductuales. Esta es la razón por la que, generalmente, no suelen sistematizarse.

Por último, es preciso añadir que la «Aleatorización» constituye, sin duda, una de las técnicas de control más ampliamente utilizadas. Cuando en un experimento se hace imposible el uso de las técnicas anteriormente descritas, se puede recurrir a la aleatorización. En virtud de la aleatorización, los diferentes valores de la variable de confundido se distribuyen al azar dentro de cada uno de los grupos experimentales, con lo que su posible efecto queda neutralizado. No obstante, como indican Meyers y Grossen (1974, p. 133), la aleatorización no es un procedimiento recomendable —aunque muchas veces sea la única técnica disponible—, dado que con él se aumentan las diferencias individuales ya esperadas en todo experimento.

3.1.2. *Variables de confundido de tarea*

Las variables de tarea se refieren tanto a los estímulos y materiales que son presentados a los sujetos como a las instrucciones que les son dadas a fin de que ejecuten una determinada actividad. Es decir, cualquier tarea ejecutada por el sujeto a fin de generar datos experimentales depende, fundamentalmente, del tipo de aparato o técnica de presentación de estímulos, de la propia modalidad o característica de los estímulos, así como de la serie de instrucciones o consignas dadas para su adecuada ejecución.

Es evidente, por otra parte, que las características físicas o materiales de los estímulos pueden convertirse en una potencial fuente de variables de confundido. Supongamos, por ejemplo, que se pretende estudiar el efecto de la «complejidad estimulativa» sobre la «conducta exploratoria». A tal propósito se construyen dos clases de pautas estimulativas: Bandas negras horizontales sobre un fondo blanco y bandas negras verticales. Si la variable independiente queda definida en función del sentido que toman las bandas negras, un problema que sin duda es posible que se plantee es el relativo a la amplitud de dichas bandas. Si dicha variable no queda controlada podría convertirse en una posible fuente de confundido para los resultados de este experimento. Tenemos, por tanto, en este caso una posible variable de confundido de tarea. Son frecuentes, también, las variables de confundido de tarea dentro del campo de la memoria y psicofísica. Si por ejemplo, en psicofísica nos interesase estudiar el umbral de la tonalidad, comprobaríamos que cualquier cambio en la intensidad del estímulo auditivo podría confundir con los cambios de frecuencia del mismo.

Cuando por tarea se entiende la actividad que debe ejecutar el sujeto, deberá asegurarse que ésta presente un mismo nivel de dificultad para los diferentes grupos experimentales, de forma que la característica propia de cada tarea no favorezca más a un grupo que a otro.

Un sistema frecuentemente utilizado, en relación a la variable de tarea, se ejerce eliminando aquellos sujetos que no dan, en una serie de pruebas de entrenamiento un determinado nivel o no alcanza el criterio establecido.

Estas y otras técnicas, como las que hemos descrito en el apartado anterior pueden ser utilizadas adecuadamente para el correcto control de las variables de tarea, cuando éstas pueden introducir una distorsión sistemática en los resultados.

3.1.3. *Variables de confundido de procedimiento*

Un problema relacionado con el procedimiento experimental se refiere a la forma cómo los materiales o estímulos de tarea son presentados a los diferentes grupos experimentales. Cuando se utilizan diferentes sistemas de presentación es posible que se produzcan variaciones inadvertidas aunque sistemáticas en las respuestas de los sujetos. Una consecuencia de ello es que en las respuestas se manifiesten tendencias que reflejan la especial forma de presentación de los estímulos. Una de estas tendencias es conocida con el nombre de «preferencia de la posición». Se da «preferencia de la posición» cuando el sujeto reacciona más bien a la localización, ya sea temporal o espacial, del estímulo que a sus particularidades específicas. Así, por ejemplo, si se presentan visualmente dos estímulos, X e Y, de tal forma que a X se le asocia una recompensa y a Y no y, por otra parte, que X aparezca siempre a la derecha e Y a la izquierda, es posible que nos preguntemos si el sujeto va a aprender a discriminar en función de la localización o en función de las características diferenciales de los estímulos visuales. Es posible, también, que el sujeto responda a ambas cosas a la vez. Tenemos por tanto, en este caso una posible fuente de confundido que depende de la presentación espacial de los estímulos.

Un segundo efecto de procedimiento puede darse cuando la presentación de los tratamientos sigue, invariablemente, una secuencia u orden temporal. De esta forma, si los diferentes tratamientos son aplicados a los respectivos grupos experimentales coincidiendo con momentos temporales distintos (hora del día, día de la semana, etc.) y claramente diferenciados, puede producirse un fenómeno de confundido. En este caso el tiempo puede actuar como una segunda variable coincidiendo con los diferentes valores de la variable manipulada. Como afirman Borkowski y Anderson (1977, p. 133), «la localización o situación espacial y temporal de los materiales de estímulo o tratamientos representan fuentes de confundido si tales elementos covarían inadvertidamente con la variable independiente». En el caso que se dé esta presunta covariación, se hace totalmente imposible determinar hasta qué punto los cambios creados en la variable dependiente es efecto de los tratamientos experimentales o bien del factor de covariación.

Un tercer fenómeno de confundido, en relación a la disposición de los tratamientos, suele aparecer en aquellas situaciones en que cada uno de éstos se aplica, en su totalidad, a todos los sujetos del grupo experimental. Es evidente que dicho efecto depende, en gran parte de la estructura propia del

diseño experimental, sobre todo de aquellos diseños conocidos con el nombre de «diseños intra-sujetos». Y aunque más adelante hablaremos de estos efectos con más detalle al tratar el tema de los diseños, vamos a señalar, sucintamente que tales efectos, conocidos generalmente como «efectos secuenciales», «efectos de orden», «efectos transferenciales» y «efectos residuales acumulativos» son propios de los diseños intra-sujeto y se producen cuando la acción de un tratamiento persiste en el tiempo de forma que puede influir sobre la posible acción ejercida de un segundo tratamiento. De esta forma, el efecto actual del segundo tratamiento sobre la variable dependiente queda mediatizado por la presencia de un efecto residual o acumulativo del primer tratamiento.

Un cuarto efecto puede producirse, también, en aquellos diseños experimentales en los que se exige del sujeto una respuesta repetitiva (como en los experimentos de aprendizaje). En tales casos, es frecuente que se generen efectos secundarios que tiendan a mejorar o a deteriorar la ejecución. Como afirman Borkowski y Anderson (1977, p. 134), la respuesta repetitiva puede afectar la ejecución del sujeto en dos sentidos, «en primer lugar, puede mejorar la ejecución a causa de factores tales como la creciente familiarización con los materiales y la exigencia de la tarea, el desarrollo de una disposición a responder o a aprender, y una creciente confianza debida a los primeros éxitos en la tarea». Estos mejoramientos de la tarea son conocidos, también, con el nombre de «efecto de la práctica». Tales efectos de práctica, que se traducen en un gradual mejoramiento en la ejecución, se deben al «calentamiento» o «warm-up» y a un componente más permanente denominado «aprendizaje a aprender» (Thune, 1950). También, el efecto de la respuesta repetitiva del sujeto puede traducirse en un deterioro de la ejecución debido a factores tales como la fatiga, habituación y adaptación. Todos estos efectos reciben el nombre genérico de «efectos de fatiga».

Una última variable de confundido relacionada con el procedimiento es la relativa a las instrucciones o consignas dadas a los sujetos para la correcta ejecución de una determinada actividad. Gran parte de las tareas realizadas por los sujetos depende, fundamentalmente, de la naturaleza y carácter de las instrucciones leídas. Por todo ello es evidente que dicha variable se puede convertir en una poderosa fuente de confundido. Esta es la causa por la que en el momento de redactar las instrucciones se dé la necesaria información al sujeto para que pueda ejecutar la tarea, sin que por ella pueda enterarse del objetivo básico del experimento. Las instrucciones deben posibilitar, como norma general, que la variable independiente ejerza adecuadamente su acción sobre la variable de medida, sin que se conviertan en causa de la misma. Cabe señalar, también, que en algunos experimentos las instrucciones se convierten en variable independiente. Tenemos, por ejemplo, aquellos casos en que a través de las instrucciones se pretende crear una actitud o inducir una creencia a los sujetos. De este modo, mediante unas instrucciones dadas diferencialmente a los diversos grupos experimentales se pretende desarrollar distintas

disposiciones subjetivas capaces de afectar cualquier fenómeno de tipo perceptivo o estimativo. Cuando esto ocurre cabe plantearnos la siguiente cuestión: ¿Cuál es la verdadera causa de los cambios operados en la conducta? ¿Las instrucciones entendidas en un sentido estrictamente físico, o bien, la disposición o actitud suscitada a los sujetos, o ambas cosas a la vez?

Algunos investigadores están de acuerdo en afirmar que las diferentes disposiciones o actitudes inducidas a través de las instrucciones constituyen la única causa o variable explicativa de los hechos. No obstante, un análisis más profundo de dicha variable nos hace caer en la cuenta de que está integrada por dos componentes: (a) La naturaleza física de un enunciado y (b) cualquier tendencia o disposición interna inducida por la misma. Estos dos componentes suelen combinarse en algún sentido, influyendo, así, en la respuesta verbal del individuo.

Uno de los presupuestos básicos que establecen Borkowski y Anderson (1977), cuando se refieren a la utilización de las instrucciones es que se conviertan «simplemente en un modo de inducir una respuesta hipotética interna que, a su vez, cause la ejecución en cuestión» (p. 124). Es evidente que dentro de este contexto, las instrucciones son entendidas o interpretadas desde una perspectiva mediacional. Es decir, las instrucciones pueden ser consideradas, inicialmente, como una clase de variable experimental física que suscita en el sujeto un tipo de respuesta interna —disposición o actitud—, la cual a su vez es causa de una respuesta verbal. De esta forma, las instrucciones pueden ser consideradas como estímulos físicos que generan en los individuos «respuestas o reacciones internas», que a su vez influyen en sus respuestas verbales. Un análisis más detallado de la variable «instrucciones» nos lleva a concluir que éstas se hallan integradas por un conjunto de estímulos verbales, cuya variación puede afectar sensiblemente las respuestas internas o mediacionales de los sujetos. Por dicha razón hay que tener sumo cuidado en su redactado, de forma que sólo susciten aquellas respuestas que sean requeridas para los propósitos básicos del experimento. En efecto, cualquier cambio de una palabra puede tener el mismo efecto sobre los resultados que la alteración de una pauta estimulativa o aumento de la intensidad de un estímulo físico. También puede ocurrir que las instrucciones sean entendidas de diferente forma por los sujetos, como también, que la introducción de una o más palabras puede alterar el significado total de una frase. De todas estas consideraciones se deriva la gran importancia que tienen las instrucciones en la investigación experimental, en especial cuando son utilizadas como variable manipulada.

Como colofón al análisis realizado sobre las instrucciones, como variable de confundido, vamos a indicar unos criterios que permitan restringir al máximo su efecto como variable secundaria: (1) Pueden ser leídas de nuevo a fin de que sean totalmente entendidas por los sujetos; (2) Podemos pedir a los sujetos que nos las expliquen a fin de comprobar que han sido perfectamente comprendidas: (3) Se pueden utilizar ensayos de preentrenamiento para verificar si efectivamente han sido comprendidas y (4) Se pueden manipular con

objeto de determinar si realmente poseen la influencia que realmente se espera de ellas.

3.2. Variables de confundido de sujeto

En todo experimento se hallan presentes una serie de variables de sujeto (Underwood y Shaughnessy, 1978, p. 30) conocidas con el nombre de diferencias interindividuales que pueden, en un momento dado, confundir la acción de los tratamientos. Por lo general dichas variables recogen las características biológicas, sociológicas y psicológicas relativamente estables de los sujetos, tales como la edad, sexo, rasgos de personalidad, capacidad intelectual, intereses, valores, etc. Estos aspectos que reflejan dimensiones estables de los sujetos pueden tener una acción de confundido cuando covarían sistemáticamente con los valores o niveles de la variable independiente. No obstante, dichas variables suelen quedar aleatorizadas, mediante el sistema de muestreo, pasando a formar parte de lo que se conoce con el nombre de «error experimental». En el caso que sistemáticamente estuviesen presentes en un determinado grupo experimental, se produciría un error muestral, o «error de tipo S» de acuerdo con la nomenclatura propuesta por Lindquist (1953).

No obstante, cuando se consideran las variables de sujeto, una de las principales fuentes de confundido procede del hecho de que los sujetos que participan en el experimento sean sujetos voluntarios o pagados, o bien sujetos que hayan participado previamente en experimentos de laboratorio. Como señalan Rosenthal y Rosnow (1975, p. 85), «existe una considerable evidencia que sugiere que, dentro del contexto de nuestra cultura occidental, el papel de sujeto de investigación es probablemente bien conocido por la mayoría de sujetos adultos que encuentran su forma en nuestras poblaciones de sujetos». En Psicología, concretamente, la mayoría de sujetos proceden de poblaciones de estudiantes que se hallan dispuestos a colaborar (Higbee y Wells, 1972). Como ha subrayado McNemar (1946), «la actual ciencia de la conducta humana es ampliamente la ciencia de la conducta de los estudiantes universitarios».

Rosnow y Aiken (1972), partiendo del supuesto de que la experimentación constituye una situación de carácter social, han propuesto una interpretación del «papel del sujeto de investigación» basados en un modelo de procesamiento de información de cinco pasos o estadios.

El primer paso, que constituye el origen de la investigación, está formado por las expectaciones que tienen los sujetos «como sujetos de investigación». Todo experimento se halla integrado por una serie de elementos que determinan unas expectativas concretas en el sujeto en relación al mismo. Estos aspectos que se desarrollan en el sujeto en función de su autopercepción como sujeto de investigación y en función de las hipótesis del investigador son denominados por Orne (1962) «características de la demanda». Las características de la demanda, como hemos indicado, abarcan todas aque-

llas sugerencias que el sujeto adquiere no sólo sobre su actual papel como sujeto experimental, sino también en base a los propósitos de la investigación. Las «características de la demanda» son evaluadas por el sujeto en función de la información paralela que obtiene acerca del experimento, como las instrucciones, contexto, filtraciones que recibe sobre el objetivo de la investigación y, sobre todo, de las impresiones que recibe el experimentador. Como ha señalado insistentemente Rosenthal (1966, 1967, 1969), una importante fuente de las características de la demanda son las preconcepciones que posee el sujeto, así como el modelamiento de su conducta. El experimentador puede modelar la conducta del sujeto en la medida que proyecte sobre él su especial concepción y particular expectativa que tiene del experimento. En efecto, el tono emocional que adopta el experimentador en torno a una ejecución puede convertirse en fuente de aprobación o reprobación para el sujeto. De esta forma se modela, indirectamente, la conducta del sujeto de acuerdo con el paradigma de acción de los reformadores sociales. En este primer paso, pues, el sujeto adquiere «las expectaciones que se tienen como sujeto de investigación», a pesar de que tales expectaciones pueden quedar alteradas por otros factores internos del sujeto. Por ello, como afirman Rosenthal y Rosnow (1975, p. 85), «con objeto de que estas características de la demanda puedan ser procesadas deben ser, en primer lugar, adecuadamente recibidas».

El segundo paso del modelo propuesto por Rosnow y Aiken (1973) se refiere a cómo el sujeto interpreta y reconoce las características de la demanda. Esto dependerá de la capacidad y habilidad del sujeto en captar su importancia como elemento decisivo en la investigación que se lleva a cabo. El típico sujeto de investigación es una persona especialmente sensible y dispuesto a cooperar activamente con los propósitos del investigador, hecho que le convierte en un ser especialmente receptivo a cualquier tipo de señal que pueda orientar su tarea en la situación exprimental. Se ha podido comprobar que los sujetos que son especialmente receptivos a las características de la demanda favorecen positivamente la comprobación de la hipótesis.

El tercer paso se halla relacionado con la motivación del sujeto o capacidad de orientar su conducta de acuerdo con sus percepciones. Evidentemente los sujetos que voluntariamente han aceptado el papel de sujeto de investigación se muestran más adquiescentes en relación a las características de la demanda. Como han señalado Rosenthal y Rosnow (1974), los sujetos voluntarios tienden a ser más complacientes a tales características que los sujetos no voluntarios. Este hecho es interpretado por Rosenthal y Rosnow (1975) en el sentido de que la cooperación activa del sujeto con el investigador puede ser considerada como una autojustificación, o como resultado de una autoselección o quizá, también, por el hecho de que el sujeto voluntario suele presentar unos rasgos psicológicos típicamente específicos (gente educada, de un mejor estatus social, que busca una mayor aprobación, menos autoritaria, más complaciente, más sociable, más joven, etc.). Puede darse, como es lógico, situaciones totalmente opuestas. Sobre todo cuando las ca-

racterísticas de la demanda se hallan en total contradicción con una evaluación desfavorable en relación a un atributo psicológico importante. Cuando se da dicha situación, el sujeto en lugar de cooperar y mostrarse complaciente con las características de la demanda, puede reaccionar en forma totalmente contraria. Existen, por tanto, una serie de factores que pueden, en un momento dado, suscitar una disposición motivacional contraria a los propósitos de la investigación.

Rosnow y Aiken sitúan en el cuarto paso de su modelo la capacidad del sujeto de responder positivamente a sus motivaciones. Existen, sin duda, tareas que requieren por parte del sujeto la ejecución de un tipo de actividad para la que no se halla biológica ni físicamente dotado. En este caso, el sujeto más bien dispuesto y más ávido en colaborar se ve incapaz de cumplir con las características de la demanda. Evidentemente, este estadio intermedio del modelo del papel del sujeto de investigación depende básicamente del experimentador. El experimentador no va a exigir del sujeto la ejecución de una actividad que se halle fuera de los límites de sus capacidades normales. Esta es la razón por la que Rosnow y Aiken consideran como pasos más importantes, en relación al sujeto, la receptividad y la motivación (2 y 3).

El último paso describe la conducta del sujeto que se halla operativamente bajo la influencia básica de los pasos 2 y 3. La conducta final recoge, fundamentalmente, el efecto combinatorio de estos dos pasos mediadores y en función de ellos pueden darse tres estados mutuamente exclusivos y exhaustivos: Adquiescente con las características de la demanda (A^+) contra-adquiescente (A^-) y no adquiescente (A^0). Así, la conducta final del sujeto puede ser considerada como el resultado final de los pasos mediadores 2 y 3, junto con el paso 4.

La siguiente figura presenta diagramáticamente la secuencia de pasos intervinientes del modelo descrito y permite una versificación visual de los efectos de sesgo que puede producir la reactividad observacional, cuando los sujetos adquieren el papel de «sujetos de investigación».

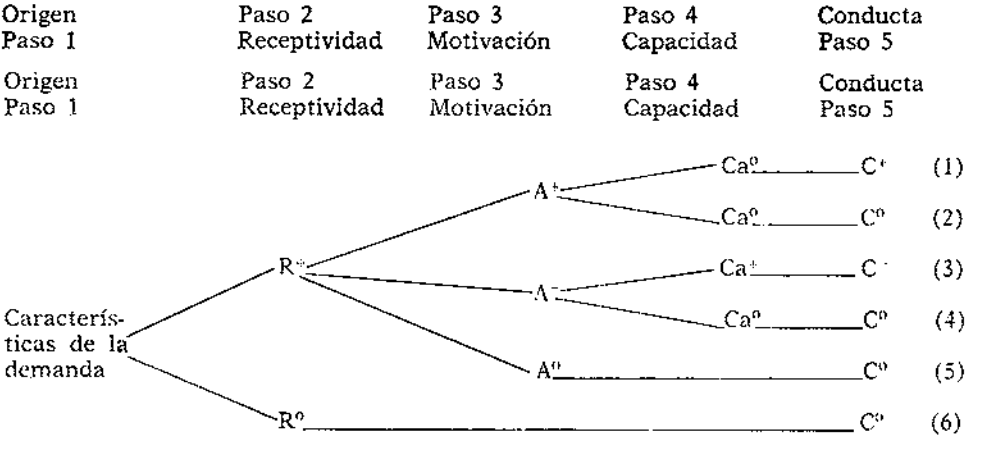


FIG. 1. Modelo de procesamiento de información propuesto por Rosnow y Aiken (1973).

En este diagrama en árbol se presentan seis ramas que, a excepción de la 1 y 3, llevan a una conducta que debería verse libre del error reactivo del sujeto. La primera vía lleva a un tipo de conducta propia del sujeto que «adopta un papel de sujeto de investigación». Es adquiescente con la demanda, percibe los deseos del investigador, está positivamente motivado y es capaz de desempeñar un buen papel como sujeto de investigación. La vía 3, presenta una conducta de un sujeto motivado de forma negativa. Las vías 2 y 4, se hallan limitadas por la falta de capacidad, la 5 por la falta de motivación y la 6 por la falta de receptividad. Desde el punto de vista metodológico el principal problema consiste en salvar, para las vías 1 y 3, la posible reactividad observacional que posee un efecto de desviación sistemática sobre los resultados.

Se han propuesto algunas alternativas con el fin de restringir o eliminar la «reactividad observacional» que se produce en las situaciones 1 y 3, siendo de todas ellas, quizá, la más plausible la propuesta por Campbell (1969), quien sugirió la posibilidad de planificar los experimentos en situaciones naturales, fuera del ambiente de laboratorio. «Si el problema de la reactividad surge del hecho de que los sujetos generan conductas artificiales en respuesta a las características de la demanda cuando consideran que un investigador les está observando, una solución alternativa es desviar la atención del sujeto de las características de la demanda y orientarles hacia demandas sociales naturales» (Rosenthal y Rosnow, 1975, p. 89).

3.3. Variables de confundido del experimentador

Una de las variables que pueden distorsionar los resultados de un experimento es el «experimentador en sí mismo». Es evidente, aunque ello nos sorprenda, que cualquier cosa que conozcamos sobre el experimentador como persona y su actuación con respecto al sujeto de investigación es algo muy importante para la ciencia experimental. No vamos aquí a introducir la distinción que Barber (1973, p. 382) establece entre «investigador y experimentador», en su estudio sobre los nueve efectos de investigador y experimentador. Nos vamos a limitar a los efectos de experimentador, tal como tradicionalmente ha sido propuesto por Rosenthal (1966), Friedman (1967) y Rosenthal y Rosnow (1969).

Como Borkowski y Anderson (1977) indican, la variable experimentador constituye un elemento extraño, de gran importancia, que puede llegar a confundir la acción de la variable independiente y, como consecuencia, distorsionar los resultados de una investigación. En tal caso se obtendría un diseño totalmente contaminado atentándose abiertamente contra su validez interna de forma que toda conclusión derivada del mismo carecería de valor. De ello se deriva la importancia que tiene el estudio del «efecto de experimentador», ampliamente estudiado por Rosenthal (1966) en Harvard y posteriormente por Rosenthal y Rosnow (1969) y Barber (1973).

Rosenthal (1966) en la descripción inicial que hace sobre el «efecto de experimentador» introduce una distinción básica, que por otra parte, según Barber (1973), debería ser atribuida distintamente al «investigador» de un lado y al «experimentador» de otro. A pesar de que Rosenthal no insiste en dicha distinción, los efectos que indiscriminadamente atribuye al experimentador recogen, básicamente, los principales errores que pueden cometer tanto el investigador como el experimentador. Estos efectos son: (a) la influencia que el experimentador ejerce directamente sobre los datos del experimento, independientemente del sujeto y (b) el efecto de experimentador, propiamente dicho el cual se produce cuando el experimentador, como persona, interactúa con el sujeto proyectando sobre él sus expectativas, su forma de ser, sus actitudes, etc., todo lo cual puede producir un sesgo importante en la conducta del sujeto de investigación.

Uno de los riesgos al que el experimentador se expone en relación a los resultados recibe el nombre de «error de registro». Este error puede darse en aquellos casos en que el experimentador debe observar la conducta de los sujetos y registrarla de alguna forma. Se ha podido constatar la existencia de errores de registro, cosa frecuente en cualquier investigación y que, por otra parte, dichos errores no han sido compensados entre sí. Un grupo de errores compensados sería aquel en que la distribución del error es al azar y no favorece ningún tipo de respuesta o medida de la variable dependiente.

Los errores de observador deben distinguirse del «sesgo del observador», dado que, por lo general, los primeros se distribuyen al azar en torno al criterio o valor correcto. Las observaciones sesgadas suelen presentarse consistentemente desviadas en un sentido u otro y pueden reflejar alguna relación con respecto a las características del observador o de la situación observada, o de ambas cosas a la vez.

Al considerar el acto o secuencia de actos que constituye la observación en el quehacer científico se puede establecer una distinción, a nivel conceptual, entre localización del «error de registro» y «sesgo». El sesgo, conocido también con el nombre de error de captación, suele depender de las diferentes posiciones del experimentador, de los ángulos de la observación, de las imperfecciones del aparato sensorial, de los sistemas de conexión nerviosa, de las áreas de proyección cortical, etc. El error de registro, por su parte se distingue del anterior, ya que puede darse partiendo de la base de la no existencia de un error de captación, en la transcripción de los datos o la respuesta en el sistema de registro, lo cual difiere completamente del hecho que sea correctamente captado. Es evidente que una de las posibles técnicas de control de estos dos errores (registro y captación) consiste en la replicación del experimento. Con frecuencia la replicación de las observaciones sirve para establecer el carácter de este tipo de errores.

Junto a estos efectos que inciden directamente sobre el experimento, en el registro de datos, análisis de los mismos, interpretación, planificación, etc.,

Rosenthal (1966) dedica gran parte de su trabajo al estudio del efecto propiamente de experimentador o la interacción entre éste y la respuesta de los sujetos. Rosenthal en su libro trata en profundidad del análisis de los siguientes efectos de experimentador:

1. Interacción debida a los atributos biosociales (sexo, edad, raza).
2. Interacción debida a los atributos psicológicos (ansiedad, hostilidad, autoritarismo, inteligencia).
3. Interacción debida a los factores situacionales (familiaridad, experiencias en la conducción de experimentos, etc.).
4. Interacción debida al modelamiento (grado de correlación entre la ejecución del experimentador y del sujeto en la misma tarea).
5. Interacción debida a la espectativa del experimentador sobre los resultados del experimento.

Dicha clasificación no es exhaustiva y la podríamos completar con la que nos propone Barber (1973, p. 383), quien, como ya hemos indicado, establece una distinción inicial entre investigador y experimentador. Según Barber, el investigador es quien decide qué tipo de trabajo o estudio ha de realizarse, cómo debe ser diseñado, cómo debe ser analizado o interpretado. El investigador, por su parte, es responsable del experimento, en su totalidad, así como del previo entrenamiento que han de recibir los experimentadores para su correcta ejecución. El experimentador, de otra parte, es quien lleva a la práctica el trabajo, probando a los sujetos, administrando las instrucciones y registrando sus respuestas. De ahí que mientras el investigador es el responsable del diseño experimental, el experimentador es responsable de la recogida de datos. A partir pues de esta distinción es obvio que puedan establecerse efectos específicos para cada uno de estos dos componentes. Es decir, los efectos de investigador y los efectos de experimentador. De acuerdo, pues, con Barber, estos efectos pueden describirse de acuerdo con el siguiente esquema:

Efectos de investigador

1. Paradigma del investigador.
2. Protocolo vago o impreciso.
3. Análisis del investigador.
4. Falseamiento intencional de los resultados.

Efectos de experimentador

5. Atributos del experimentador.
6. Fracaso en seguir el protocolo.
7. Registro erróneo de los datos.
8. Efecto de falseo intencional.
9. Las expectativas no intencionales.

Estos son, en definitiva, los principales efectos que pueden producirse en una investigación en función tanto del investigador como del experimentador. Esta clasificación recoge, fundamentalmente, la establecida por Rosenthal en 1966.

4. DISEÑO EXPERIMENTAL

Según Meyers y Grossen (1974, p. 140), el término «diseño experimental» o simplemente «diseño», en el contexto de investigación, puede significar lo siguiente:

1. El término diseño entendido como «verbo» viene a significar un proceso conjunto de actividades encaminadas a la planificación de un experimento. Así, diseñar un experimento, significa tomar una serie de decisiones en torno a los grupos experimentales, a los valores de las variables (independiente y dependiente), aplicación de controles, utilización de pruebas estadísticas, etc.
2. En segundo lugar, el término diseño es utilizado para describir «una clase específica de experimento». Dentro de esta segunda acepción se suele utilizar términos tales como «diseño de transferencia», «diseño de condicionamiento», «diseño de memoria a corto plazo», etc. En todos estos casos, el diseño se refiere a un tipo concreto de procedimiento utilizado para formar los respectivos grupos experimentales con objeto de probar un determinado efecto.
3. Se entiende, también, por diseño a la forma como específicamente van a disponerse los diferentes grupos que formarán parte de un experimento. De acuerdo, pues, con este criterio, los diseños suelen conocerse como diseños de dos grupos separados, diseños multigrupos, diseño de un sólo grupo, diseño factorial, etc.
4. Por último, una cuarta utilización del término diseño hace referencia al tipo de análisis estadísticos que van a aplicarse a los datos. Así, tenemos los diseños paramétricos, diseños no paramétricos, según la prueba de significación utilizada se ajuste o no a la teoría normal.

De entre estas cuatro acepciones quizá la más utilizada, en la mayoría de los tratados sobre el diseño experimental, sea la tercera. Es decir, el diseño puede ser considerado en función de la cantidad y forma como se han constituido los grupos experimentales. Éste va a ser el concepto que vamos a utilizar en el presente trabajo. Por otra parte, esta particular concepción del diseño, si bien no excluye a las otras, nos va a servir para una mejor discusión de los temas que hemos introducido anteriormente: validez y control.

Uno de los procedimientos para conseguir un control efectivo de las variables de confundido, que pueden de una forma u otra atentar contra la validez de una investigación, es mediante la utilización de un adecuado diseño experimental. En todo diseño deben conjugarse las condiciones básicas que aseguren tanto la validez interna como la externa. Cada diseño, requiere, por su parte, unos determinados sistemas de control de las variables de confundido, así como unos procedimientos específicos para la formación de los diferentes grupos de tratamiento. Mediante el control quedará garantizada la validez interna del diseño y mediante la apropiada selección de la muestra para la formación de los grupos, se garantizará su validez externa.

Un diseño proporciona una relativa validez externa, en la medida que los sujetos de la muestra representen proporcionalmente a la población sobre la que van a extenderse los resultados. De ahí, que la validez externa depende, en gran parte, de la forma o método de seleccionar los sujetos para la formación de los diferentes grupos de tratamiento. Ésta constituye una de las razones para que al considerar el diseño experimental no sólo se tenga en cuenta la forma como se controlan las variables extrañas, sino también el procedimiento mediante el cual se garantice su validez externa. En la descripción que a continuación vamos a hacer de los diseños de experimentación nos fijaremos, fundamentalmente, en los dos siguientes aspectos: (a) cómo a través del diseño se podrá conseguir un adecuado control de las variables de confundido, y (b) cuáles van a ser las estrategias que se utilizan, en cada diseño, para una efectiva formación de los diferentes grupos experimentales. Sólo en base a estos dos criterios se podrá valorar al diseño desde una perspectiva total en cuanto a la validez.

4.1. Clasificación de los diseños experimentales

Los diseños más ampliamente utilizados en la investigación experimental son fundamentalmente dos: (a) Diseño entre sujetos, conocido también con el nombre de grupos separados o diseño entre-grupos, y (b) Diseño intra-sujetos o diseño de un solo grupo, cuya característica más importante es el hecho de que se aplica a cada uno de los sujetos del grupo todos los tratamientos experimentales. Es evidente, por otra parte, que estas dos clases básicas de diseño poseen sus limitaciones y ventajas, y su aplicación a un

caso concreto debe valorarse en función de las condiciones experimentales así como de la disponibilidad de los sujetos.

4.1.1. *Diseño entre-sujetos o de grupos separados*

Uno de los principales riesgos que poseen los diseños entre grupos, en relación a la validez interna, es que la diferencia constatada entre los diferentes grupos no sea causada por el tratamiento experimental, sino por las diferencias de los sujetos que pertenecen a los respectivos grupos. Una tal diferencia de sujetos en los grupos, previa a la aplicación de los tratamientos, puede ejercer un efecto contaminante sobre los datos y falsear totalmente su interpretación. Si los grupos están formados de forma que se produzca una diferencia consistente entre ellos, la variable dependiente queda confundida sistemáticamente con la variable de sujeto. Con lo que el experimento quedaría invalidado y no serviría como instrumento probatorio de la hipótesis experimental.

Una situación ideal para los diseños entre-grupos sería aquella en la que todos los miembros de los grupos fueran idénticos. De esta forma toda diferencia que se produjera entre los individuos de los grupos podría atribuirse a los tratamientos. Sin embargo, este ideal de experimentación es totalmente imposible de alcanzar dado que los sujetos suelen presentar diferencias en relación a una serie de atributos. Ésta es la razón por la que deberá sustituirse dicho ideal por un principio general consistente en que los grupos sean equivalentes antes de la aplicación de las diferentes condiciones experimentales. A fin de conseguir el cumplimiento de este principio básico la metodología experimental ha desarrollado algunas estrategias, tendentes todas ellas a este objetivo común. Intentaremos, pues, analizar dichas estrategias, al tiempo que comprobaremos cómo cada una de ellas reduce, en alguna medida, el riesgo de que el diseño carezca de validez. A efectos de simplicidad vamos a presentar el «diseño de dos grupos» como prototipo de la clase de diseños de grupos separados.

La forma más elemental de «diseño de dos grupos» es aquella en la que un grupo recibe una determinada cantidad de la variable manipulada y el otro, cero cantidad de la misma. El resto de condiciones se mantienen iguales para los dos grupos.

TABLA 1

REPRESENTACIÓN DE UN DISEÑO CLÁSICO DE DOS GRUPOS SEPARADOS

Grupos	*Tratamiento*	*Otras variables*	*Prueba de los efectos del tratamiento*
Experimental	X_1	Presentes	O_1
Control	X_2	Presentes	O_2

De acuerdo con este esquema, el grupo experimental recibe una determinada cantidad de tratamiento, mientras que el grupo control no recibe ninguna cantidad. Después que ambos grupos han sido sometidos a tales condiciones son probados en relación a alguna tarea a fin de obtener diferencias de ejecución. Este diseño clásico de dos grupos, permite al experimentador tomar una decisión sobre si el tratamiento tuvo algún efecto sobre la conducta de los sujetos del grupo experimental en comparación a la conducta de los sujetos del grupo control.

Una variación del diseño clásico de dos grupos viene representada en la tabla 2. De acuerdo con dicho esquema, el grupo control también recibe el tratamiento experimental, pero a un nivel menor que el grupo experimental. Es evidente que este diseño experimental es útil en algunas situaciones de investigación en las que es necesario comparar dos niveles de actuación de la variable manipulada.

TABLA 2

ESQUEMA DE UNA VARIACIÓN DEL DISEÑO CLÁSICO DE DOS GRUPOS

Grupos	*Tratamiento*	*Otras variables*	*Prueba de los efectos del tratamiento*
Experimental	X_1	Presentes	O_1
Control	X_2	Presentes	O_2

Dado que en ambos diseños se introduce un grupo de control con características similares al grupo experimental, a excepción de la cantidad de tratamiento que reciben, puede esperarse que las principales fuentes de confundido quedan totalmente neutralizadas. Estos dos diseños, considerados también como estrategias de control, pueden ser combinados para producir un diseño mucho más potente, en el que se emplean dos grupos de control de forma que (a) el primero nos va a permitir comprobar si un determinado fenómeno experimental ha ocurrido o no (comparación del grupo experimental con el grupo de control 1), y (b) la total ausencia de tratamiento del segundo servirá para comprobar si existe algún efecto inesperado o no pretendido en la variable de medida (comparación del grupo de control 1 con el grupo de control 2).

TABLA 3

ESQUEMA DE UNA VARIACIÓN DEL DISEÑO CLÁSICO DE DOS GRUPOS

Grupos	*Tratamiento*	*Otras variables*	*Prueba del efecto del tratamiento*
Experimental	X_1	Presentes	O_1
Control 1	X_2	Presentes	O_2
Control 2	X_0	Presentes	O_3

Dos consideraciones básicas en torno a este tipo de diseños, en relación a la validez tanto externa como interna se refieren a: 1) cómo los sujetos han sido seleccionados, y 2) cómo los sujetos seleccionados han sido asignados a los diferentes grupos experimentales. Según el modo como el diseño resuelva estas dos cuestiones primordiales se podrá establecer el grado en que el diseño garantiza tanto la validez externa como interna. Mediante una adecuada selección se puede esperar, que la muestra seleccionada, represente adecuadamente a la población de origen. Mediante una adecuada estrategia de asignación podremos valorar el grado en que los diferentes grupos pueden ser considerados equivalentes. Si un adecuado muestreo o técnica de selección es la condición básica de la validez externa, una pertinente estrategia de asignación es la condición básica de la validez interna. De acuerdo con este criterio describiremos a continuación las principales estrategias que se pueden utilizar, tanto para la selección de la muestra como para la formación de los grupos.

4.1.1.a. *La aleatorización*

La aleatorización puede ser utilizada como técnica de selección de muestra y como técnica de formación de grupos. Como técnica de selección de muestra, ésta puede considerarse representativa en la medida que la población se halle adecuadamente representada en la misma. Ello es difícil de conseguir, sobre todo, cuando las poblaciones de origen son muy heterogéneas, a no ser que se aumente notablemente el número de sujetos que van a constituir la muestra. No obstante, si la aleatorización se ha llevado con todo rigor, se puede esperar con un alto grado de confianza que la muestra sea representativa.

Por otra parte, la aleatorización puede ser utilizada como técnica de asignación de los sujetos de la muestra a los diferentes grupos de tratamiento. Con ello se pretende conseguir la equivalencia de los mismos antes que se les apliquen los tratamientos. Si en esta segunda fase, la aleatorización se ha aplicado con extremo rigor, puede asumirse, desde un punto de vista teórico, que los grupos serán equivalentes en relación a las características de los sujetos.

Tanto en un caso como en otro, la efectividad de la técnica de la aleatorización depende, en gran manera, del tamaño de la muestra. A mayor tamaño mayor efectividad. Mediante esta técnica, siempre que se cumplan las condiciones básicas del azar, se puede garantizar, hasta cierto punto la validez externa e interna del diseño.

4.1.1.b. *Apareamiento de sujetos*

En algunas situaciones es posible obtener alguna información sobre la medida de ejecución de los sujetos antes de la aplicación de los tratamientos.

Si ello es así, se puede utilizar dicha información como base para la formación de grupos equivalentes. En efecto, a partir de estas medidas antes se puede conseguir dos grupos equivalentes en relación a alguna variable extraña que se halle directamente relacionada con la variable dependiente y que, por otra parte, puede constituir un elemento de contaminación de resultados.

La técnica de apareamiento es, pues, una técnica de «formación de grupos equivalentes». Por tanto, debe descartarse dicha técnica como procedimiento de selección. Ahora bien, si la selección de la muestra se basa única y exclusivamente en el criterio de apareo, se restringe ampliamente la validez externa del diseño. Esta reducción de la validez externa redunda en un sensible aumento de la validez interna, ya que los diseños de grupos apareados ofrecen muy buenas garantías de validez interna.

Existen, por otra parte, dos procedimientos básicos de apareamiento. El primero, denominado por D'Amato (1970), «apareo por eliminación» consiste en estimar el valor de una potencial variable de confundido antes de la aplicación del tratamiento y conseguir que la media de ambos grupos en relación a dicha variable sea la misma. De esta forma ambos grupos son equivalentes en relación a la variable de confundido y, puesto que, por otra parte, se ha eliminado su efecto diferenciador, dicha técnica recibe el nombre de «apareo por eliminación». El segundo procedimiento es descrito por D'Amato (1970) como «apareo por evaluación», consiste en manipular la variable potencial de confundido con objeto de estimar su influencia juntamente con el efecto del tratamiento. Esta última técnica implica tener en cuenta diferentes niveles de la variable potencial de confundido. Por ampliación esta segunda técnica nos lleva a la estrategia de bloqueo que recoge, básicamente, dicho principio.

4.1.1.c. *Bloques de grupos*

En síntesis, el bloqueo puede ser considerado como una técnica de apareo en el que se utiliza, en la selección de los sujetos para formar los bloques, la aleatorización. Si en la población existe una importante variable de confundido que actúa con diversos valores y en base a los cuales la divide en estratos, estos estratos o niveles pueden ser tomados como base para formar bloques a partir de cada estrato respectivamente. Dentro de cada bloque, que recoge las características básicas del estrato los sujetos son asignados al azar a cada uno de los grupos experimentales. Lo cual garantiza la equivalencia de los grupos dentro de cada bloque.

Mediante este diseño, la validez externa queda garantizada por el hecho que los sujetos son seleccionados al azar a partir de cada estrato de la población. Por otra parte, la validez interna de este diseño queda, también, altamente asegurada, dado que dentro del bloque existe menos variación que la que se produciría si los grupos hubiesen sido formados totalmente al azar.

TABLA 4

EVALUACIÓN DE LOS DISEÑOS ENTRE GRUPOS EN RELACIÓN A LA VALIDEZ INTERNA Y EXTERNA

Estrategia para la selección y asignación de los sujetos	*Validez*	*Selección*	*Asignación*
Aleatorización	Interna	Buena	Buena
	Externa	Buena	Depende de otra información
Apareamiento	Interna	—	Muy buena
	Externa	—	Depende de otra información
Bloqueo	Interna	Muy buena	Muy buena
	Externa	Muy buena	Depende de otra información

4.1.2. *Diseño intra-sujeto o de un solo grupo*

La característica fundamental de los diseños intra-sujetos es que los tratamientos son aplicados a cada uno de los sujetos del grupo. Son tan obvias las ventajas que presentan este tipo de diseños que hacen aconsejable su uso en comparación a los diseños de grupos separados: (a) se utiliza un menor número de sujetos; (b) mayor precisión experimental, y (c) mejor control de las diferencias entre las condiciones experimentales, dado que las variables de confundido extrañas actúan por igual en cada tratamiento.

Al mismo tiempo, si la muestra es una buena representación de la población, este diseño puede garantizar una buena validez externa. En cuanto a la validez interna, si bien el diseño intra-sujeto permite un mejor control de las variables de confundido extrañas, se dan unos efectos, inherentes al propio diseño, que si no se controlan bien pueden comprometer fuertemente la validez interna. Estos efectos son básicamente los siguientes: (a) efectos secuenciales debidos a la aplicación consecutiva de los tratamientos. Estos efectos son conocidos también con el nombre de efectos residuales; (b) en algunas situaciones las tareas a realizar, bajo las diferentes condiciones, no son idénticas, introduciendose un factor diferencial en cuanto a dicha variable, y por último (c) se presenta, también, un problema de tipo temporal. Es decir, los tratamientos no son aplicados al mismo tiempo, sino dentro de una secuencia temporal. Esto puede introducir cambios en función del tiempo.

Queda claro que la presencia de tales factores no afecta en modo alguno la representatividad de la muestra, ni su control aumenta la validez externa del diseño. Así mismo, un adecuado control de estas variables puede redundar en un mayor incremento de la precisión y sensibilidad del diseño. De esta

forma, el diseño intra-sujeto puede detectar el efecto de los tratamientos aún cuando su diferencia sea más pequeña. Ello convierte a estos diseños en instrumentos de trabajo experimental altamente precisos y eficaces. Por otra parte, las diferencias individuales son tratadas como una contribución constante a la variabilidad total de cada tratamiento y puede ser sustraída de cada uno de ellos en la determinación de los efectos experimentales.

Si bien, como acabamos de afirmar, en los diseños intra-sujetos se mantiene constante la variabilidad entre los sujetos a lo largo de los respectivos tratamientos, con lo que se aumenta la precisión y fiabilidad experimental, no obstante, los diferentes tratamientos no son aplicados A) «al mismo tiempo al grupo de experimentación, sino dentro de una secuencia temporal». Esto, sin duda alguna, atenta contra uno de los presupuestos básicos de la experimetación la cual exige que la aplicación de las condiciones experimentales sean «al mismo tiempo». De esta forma, la variable tiempo, puede actuar como variable de confundido y debe ser controlada con objeto de garantizar la validez interna del experimento. B) Un segundo aspecto importante que debe, también, ser controlado es el relativo a la variable de tarea. Si como hemos dicho, los tratamientos son aplicados al mismo grupo experimental, deberemos variar las tareas a no ser que queramos correr el riesgo de que los sujetos adquieran un aprendizaje (en el supuesto, claro está, de que ello no sea nuestro propósito). En estos casos deberemos construir una serie de tareas que se aparecn en cuanto a dificultad y que al mismo tiempo sean diferentes. C) Por último, un tercer efecto implícito en este tipo de experimento y quizás el más importante, es el relativo a la secuencialidad de los tratamientos. Estos efectos son también conocidos con el nombre de «efectos progresivos» o «efectos acumulativos» o «efectos residuales». Uno de los procedimientos prácticos para eliminar el «efecto progresivo» consiste en la introducción de intervalos entre tratamientos (IET) lo suficientemente amplios a fin de que se disipen totalmente los efectos residuales de los tratamientos. La utilización de IET adecuados puede eliminar los efectos residuales debidos, principalmente, a la habituación, fatiga y adaptación, factores que, como es sabido desaparecen con el tiempo. Esto, sin duda, plantea un problema relativo a la elección de un IET óptimo que disipe al máximo tales efectos. Existe poca información sobre los tiempos óptimos que deben ser empleados ya que, a veces, la introducción de estos intervalos puede alargar la tarea de recogida de datos o producir aburrimiento a los sujetos, todo lo cual crea verdaderos inconvenientes. Ésta es quizá la principal desventaja que presentan los diseños intra-sujetos cuando son comparados con los diseños de grupos separados. Queda claro, por tanto, que esta serie de secuelas debidas a la temporalidad, como a la tarea y secuencialidad, son implícitas a la propia estructura del diseño, y deben ser aplicadas adecuadas técnicas de control con objeto de que tales variables se distribuyan por igual en todos los tratamientos.

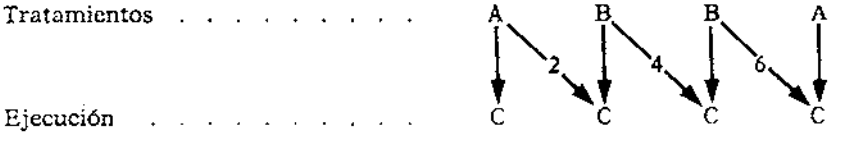
Dado que, por otra parte, las consecuencias residuales constituyen uno

de los efectos más importantes a tener en cuenta en este tipo de diseño, discutiremos las técnicas de control que permiten su neutralización; técnicas que por otra parte pueden ser hechas extensivas al resto de los efectos.

4.1.2.a. *El contrabalanceo*

El contrabalanceo constituye, de hecho, la técnica por excelencia que debe ser aplicada cuando se utilizan diseños intra-sujetos. El contrabalanceo permite distribuir proporcionalmente los efectos residuales a lo largo de las diferentes condiciones experimentales. La lógica de dicha técnica consiste en conseguir «un equilibrio» en la tasa promedio de ejecución para los diferentes tratamientos, de forma que éstos sean aplicados un número igual de veces.

Cabe añadir, además, que la técnica de contrabalanceo puede ser considerada efectiva, en la medida que el efecto residual de los tratamientos sea «uniforme y recíproco». Supongamos, por ejemplo, que estamos interesados en estudiar el efecto del «significado de las unidades verbales» sobre la «tasa de retención». Para demostrar esto construimos dos listas de triagramas. La primera, A, tiene significado, y la segunda, B, no tiene significado. El efecto de transfer o secuencial es «uniforme» si un tratamiento no adquiere más ventaja que el otro en función de su posición. Una forma de contrabalancear los tratamientos consistiría en aplicar la siguiente secuencia: ABBA. En dicha disposición experimental, el efecto secuencial sería uniforme, si hipotéticamente y a título de ejemplo, ocurriera lo siguiente: Que el tratamiento A que ocupa la primera posición no obtuviera ninguna ventaja, dado que no sigue a ningún otro tratamiento; que el primer tratamiento B, obtuviera dos puntos adicionales debido al efecto de transferencia positiva; que el segundo tratamiento B obtuviera cuatro puntos, debido a la acumulación del transfer y, que el segundo tratamiento A, recibiera un total de seis puntos adicionales como consecuencia de la acumulación de los efectos de transfer. Esto puede ser ilustrado gráficamente de la siguiente forma:



Al promediar los efectos de cada tratamiento, éstos quedarían modificados al recibir la misma cantidad de transfer; es decir, un total de seis puntos respectivamente. Con lo que el efecto de transfer puede ser considerado uniforme; es decir, actúa con el mismo valor tanto sobre el tratamiento A como sobre el tratamiento B. En el caso que no se produjera esta uniformidad del efecto acumulativo o de transfer, los tratamientos deberían ser aplicados

según un orden aleatorio o alternativo, lo cual exigiría formar más grupos experimentales.

Un segundo aspecto que hay que tener en cuenta, en relación al contrabalanceo es que los efectos residuales sean simétricos o recíprocos. El efecto residual es recíproco si la influencia del tratamiento A sobre el B es la misma que la influencia del B sobre el A. Si ello es así, el contrabalanceo se puede aplicar mediante la técnica «entre-grupos», en cuyo caso un primer grupo pasaría la secuencia AB y el segundo, BA. Si los efectos residuales tienen un carácter simétrico, no existen problemas en el momento de interpretar los resultados de los diseños contrabalanceados. El problema se plantearía en el caso que tales efectos no fueran recíprocos. Siguiendo con el ejemplo propuesto, se podría hipotetizar que el efecto de transferencia de A sobre B fuera de 5 puntos y que el de B sobre A fuera también de 5 puntos. En este caso tendríamos que el efecto residual sería simétrico. Ahora bien, si se comprueba que el efecto de transferencia de A sobre B se traduce en un mejoramiento de la ejecución en 5 puntos, y la de B sobre A, en un deterioro de la ejecución de 5 puntos, los efectos residuales no podrían ser considerados recíprocos; efectos, que por otra parte, podrían ser estimados mediante la aleatorización de los tratamientos, utilizando para ello diferentes grupos experimentales. Éstas son, en definitiva, las principales limitaciones que presentan los diseños intra-sujetos contrabalanceados y que, de producirse, atentarían gravemente contra la validez interna del diseño.

Sin duda alguna el contrabalanceo es considerado una técnica eficaz de control, puesto que los efectos residuales se distribuyen por igual, a través de las diferentes condiciones experimentales. Ahora bien, dicha técnica adquiere mayor complejidad en la medida que se aumenta el número de tratamientos. De acuerdo, por tanto, con este criterio el contrabalanceo puede ser completo, en el caso de un pequeño número de tratamientos o incompleto, cuando se dispone de mayor cantidad de tratamientos.

Por otra parte, a fin de controlar los posibles efectos debidos a la falta de uniformidad se puede utilizar la técnica del contrabalanceo entre sujetos o entre grupos. Es decir, aplicando a cada sujeto del grupo o a cada grupo una disposición diferente de los tratamientos.

a.1. *Contrabalanceo completo*

A fin de evitar los posibles efectos residuales, sobre todo, cuando éstos no son uniformes, el contrabalanceo completo entre grupos, puede ser considerado como un procedimiento ideal. En él se tienen en cuenta todas las posibles combinaciones de los tratamientos. Supóngase que tenemos tres tratamientos. Un contrabalanceo completo requeriría que se aplicasen todas las posibles combinaciones de 3 tratamientos; es decir, n! o factorial de n. En nuestro caso, tendría $3! = 6$. Lo cual quiere decir que existen seis formas

de combinar tres cosas. En tal caso, tendríamos que utilizar 6 sujetos, o seis grupos de sujetos, de forma que pudiesen ser aplicadas las seis diferentes posibles secuencias o combinaciones. Este diseño, con contrabalanceo completo entre-grupo o entre-sujetos puede ser representado de la siguiente forma:

Grupos de sujetos	*Posibles combinaciones de tres tratamientos*
Grupo 1	ABC
Grupo 2	ACB
Grupo 3	BAC
Grupo 4	BCA
Grupo 5	CAB
Grupo 6	CBA

Fijémonos, que mediante dicha técnica cada tratamiento ocurre dos veces en cada una de las tres posiciones (primera, segunda y tercera) y que cada tratamiento precede y sigue a cada uno de los otros, un igual número de veces. El contrabalanceo completo se podía haber realizado, también, con un sólo grupo de seis sujetos, es decir, mediante un contrabalanceo completo entre-sujetos. A tal efecto, los sujetos deberían haber sido asignados a cada una de las disposiciones según un criterio al azar. Esta técnica es considerada como el mejor procedimiento para el control efectivo de los efectos residuales o acumulativos propios de los diseños intrasujetos. Ahora bien, si la influencia de los efectos secuenciales no es lineal o simétrica, en el sentido que la influencia de A sobre B, es diferente que la de B sobre A, el contrabalanceo no constituye una técnica eficaz para distribuir de forma igual los efectos residuales a través de los tratamientos.

La principal limitación del contrabalanceo completo está en función del número de tratamientos. Si el número de tratamientos es superior a 4, se produce una gran cantidad de posibles combinaciones y permutaciones. Así, si se tiene 5 tratamientos, el total de disposiciones son $5 \times 4 \times 3 \times 2 \times 1 = 120$; con 6 se tiene, $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$, lo cual hace imposible llevar a cabo el experimento debido a la gran cantidad de sujetos que requiere. En cuyo caso se hace aconsejable el «contrabalanceo entre-sujetos incompleto». Es evidente, que si una de las causas que determinaban el uso del «contrabalanceo completo entre-sujetos» es debido al reducido número de sujetos que exige, cuando la cantidad de tratamientos excede a cuatro, dicho diseño raramente se aplica.

a.2. *El cuadrado latino*

La disposición de cuadrado latino puede paliar el costoso experimento que exigiría un contrabalanceo completo de una cantidad importante de trata-

mientos. La disposición de cuadrado latino, consistente en elegir al azar sólo unas cuantas disposiciones o combinaciones de tratamientos, de forma que cada uno de estos aparezca una y sólo una vez en cada fila y en cada columna. Mediante la disposición de los tratamientos en cuadrados latinos se reduce, drásticamente la cantidad de sujetos, y se consigue un efectivo control de los efectos secuenciales, en el caso que éstos sean lineales o recíprocos. Ésta sería, por otra parte, una condición básica de los cuadrados latinos en general. Siendo por otra parte, grande el número de los posibles cuadrados latinos que pueden obtenerse con diferentes tratamientos (con 6 tratamientos se obtienen 8.640 cuadrados latinos), para que el contrabalanceo sea efectivo deberá cumplirse una segunda condición básica: Que cada tratamiento siga y preceda a cada uno de los otros una sola vez. Consideremos los dos siguientes cuadrados latinos para 6 tratamientos:

Cuadrado latino

A	B	C	F	E	F
C	E	D	A	F	B
E	C	A	F	B	D
B	D	F	C	A	E
D	F	B	E	C	A
F	A	E	B	D	C

Cuadrado latino

A	B	C	D	E	F
B	D	A	F	C	E
C	A	E	B	F	D
D	F	B	E	A	C
E	C	F	A	D	B
F	E	D	C	B	A

Fijémonos que ambos cuadrados latinos cumplen con el principio general de dicha disposición: cada tratamiento aparece una sola vez en cada posición original. No obstante, el primero no cumple con la segunda condición, sí en cambio el segundo. Sólo si se dan juntas ambas condiciones, el diseño de cuadrado latino puede ser considerado como una técnica eficaz de contrabalanceo. Si, por otra parte, asignamos al azar, cada grupo o sujeto, a una de las diferentes disposiciones, el diseño de cuadrado latino puede ser considerado como una aplicación del diseño intrasujeto.

4.3. *Contrabalanceo incompleto al azar*

Paralelamente a la técnica de cuadrado latino, cabe la posibilidad, para gran cantidad de tratamientos, de utilizar el procedimiento de selección al azar de las disposiciones. Esta técnica es la menos aconsejable de las posibles formas de contrabalanceo disponibles, dado que no asegura un eficaz control de los efectos de transferencias de carácter lineal o simétrico. No obstante, también mediante dicha técnica, se puede llevar a cabo un contrabalanceo incompleto eligiendo menos disposiciones de tratamientos que las que resultarían si se tuviese en cuenta todas las posibles combinaciones. Eligiendo, pues, al azar una serie de combinaciones, del total de combinaciones posibles,

se puede contar con un procedimiento rápido para la resolución de un problema experimental.

RESUMEN

El propósito de este trabajo consiste en analizar, en profundidad, la temática de la validez experimental desde una perspectiva metodológica. En él se plantea un enfoque alternativo a la concepción clásica propuesta por Campbell y Stanley (1962). Se estudian, también, las principales fuentes de confundido que pueden llegar a comprometer la validez de los experimentos al tiempo que se describen las principales técnicas de control aplicables a cada una de éstas. El trabajo termina con una discusión sobre el diseño experimental y cómo, mediante el diseño, se puede conseguir un adecuado equilibrio entre las dos principales clases de validez en la investigación conductual.

RÉSUMÉ

Le présent travail se propose d'analyser en profondeur la question de la validité expérimentale à partir d'une perspective méthodologique. Il s'agit d'une approche alternative, face à la conception classique proposée par Campbell et Stanley (1962). Sont également étudiées les principales sources de méprise qui peuvent aller jusqu'à compromettre la validité des expériments, tout en décrivant les techniques de contrôle applicables à chacune de ces sources. L'article conclut sur une discussion concernant le plan expérimental et la façon de parvenir, grâce à ce plan, à un équilibre satisfaisant entre les deux principaux types de validité dans la recherche du comportement.

ABSTRACT

The aim of this work consist in analyzing, inprofundity, the thematic of experimental validity from a methodological perspective. In this one presents an alternative approach between the classical methods of Campbell and Ctenley (1962) and this work. At time, are profoundice the main sources of confounding that can risk the validity of experiments and at a time one describes the main techniques of control with possibility to assign to every one. The work finish with a discussion about the experimental design and what, through tis, es possible to get an adequate equilibrium between the most important kinds of behavioral validity research.

BIBLIOGRAFIA

BARBER, T. X.: «Pitfalls in research: Nine investigator and experimenter effects». En R.M.W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973, p. 382-404.

BORKOWSKI, J. B. y ANDERSON, D. C.: *Experimental Psychology*. Glenviw, Ill.: Scott, Foresman y Co., 1977.

CAMPBELL, D. T. y STANLEY, J. C.: «Experimental and quasi-experimental designs for research in teaching». En N. L. Gage (ed.), *Handbook of research on teaching*. Chicago: Rand McNally, 1962.

COLLIER, G. y MARX, M. H.: «Changes in performance as a function of shifts in the magnitude of reinforcemente». *J. exp. Psychol.*, 1959, *57*, 305-309.

D'AMATO, M. R.: *Experimental psychology: Methodology, psychophysics and learning*. Nueva York: McGraw-Hill, 1970.

DUNHAM, J. y KILPS, B.: «Shifts in magnitude of reinforcement: Confounded factors or cibtrast effects?», *J. exp. Psychol.*, 1969, *79*, 373-374.

FRIEDMAN, N.: *The social nature of psychological research*. Nueva York: Basic Books, 1967.

CAMPBELL, D. T.: «Prospective: Artifact and control». En R. Resenthal y R. L. Rosnow (Eds.), *Artifact in behavioral research*. Nueva York: Academic Press, 1969, p. 351-382.

GOTTSDANKER, R.: *Experimenting in psychology*. Englewood Cliffs, J.: Prentice Hall, 1978.

HIGBEE, K. L. y WELLS, M. G.: «Some research trends in social psychology during the 1960s». *Amer. Psychol.*, 1972, *27*, 963-966.

KEPPEL, G.: *Design and analysis: A reseracher's handbook*. Englewood Cliffs, N. J.: Prentice-Hall, 1973.

LINDQUIST, E. F.: *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin 1953.

MCNEMAR, Q.: *Psychological statistics*. Nueva York: John Wiley, 1962 (3.ª ed).

MEYERS, L. S. y GGOSSEN, N. E.: *Behavioral reserach*. San Francisco, Francisco, California: Freeman y Co., 1974.

ORNE, M. T.: «On the social psychology of the psychological experiment: With particular reference to demand characteristics and their interpretations. *Amer. Psychol.*, 1962, *17*, 776-783.

ROSENTHAL, R.: *Experimenter effects in behacioral research*. Nueva York: Appleton-Century-Crofts, 1966.

ROSENTHAL, R.: «Covert communication in thepscychological experiment». *Psychol. Bull.*, 1967, *67*, 356-367.

ROSENTHAL, R.: «Interpersonal expectations: Effects of the experimenter's hypotheses». En R. Rosenthal y R. L. Rosnow (Eds.), *Artifact in behavioral research*. Nueva York: Academic Press, 1969, p. 181-277.

ROSENTHAL, R. y ROSNOW, R. L. (Eds): *Artifact in behavioral research*, Nueva York: Academic Press, 1969.

ROSENTHAL, R. y ROSNOW, R. L.: *The volunteer subjet*, Nueva York: J. Wiley, 1974.

ROSENTHAL, R. y ROSNOW, R. L.: *Primer of methods for the behavioral sciences*. Nueva York: J. Wiley, 1975.

THUNE, L. E.: «The effect of different types of preliminary activities on subsequente learning of paired-associate material». *J. exp.* Psychol., 1950, *40*, 423-428.

ROSNOW, R. L. y AIKEN, L. S.: «Mediation of artifacts in behavioral research». *J. exp. soc. Psychol.*, 1968, *10*, 363-369.

UNDERWOOD, B. J.: *Psychological research*. Nueva York: Appleton-Century-Corfts, 1957.

UNDERWOOD, B. J. y SHAUGHESSY, J. H.: *Experimentación en Psicología*. Barcelona, Omega, 1978.

WEBB, E. J.; CAMPBELL, D. T.; SCHWARTZ, R. F., y SECHREST(L.: *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally, 1966.