

LOS DISEÑOS EXPERIMENTALES EN PSICOLOGIA

JAIME ARNAU GRAS

Laboratorio de Psicología Experimental
Departamento de Psicología
Universidad de Barcelona

1. ESTRUCTURA BÁSICA DEL DISEÑO EXPERIMENTAL

Tradicionalmente el diseño experimental ha sido concebido como una estructura lógica unitaria que, según Fisher (1935), implica dos consideraciones básicas que no son sino dos aspectos de una misma realidad. Así, pues, concretamente, para Fisher el diseño experimental supone una determinada organización de los diversos aspectos que constituyen el experimento y, a su vez, un determinado procedimiento estadístico que nos permite interpretar los resultados obtenidos. Si no se tienen en cuenta estas dos vertientes de una misma realidad lógica, podemos correr el peligro de llevar a cabo experimentos cuya finalidad inicial fuera la de «acumular datos», para recurrir inmediatamente después al estadístico a fin de «conseguir algo» (Reuchlin, 1953).

Es, pues, el diseño experimental un plan o estructura unitaria formada por una serie de prescripciones operativas referentes a la selección de los grupos experimentales, a la aplicación de los respectivos tratamientos, a la utilización de determinadas pruebas estadísticas, etc. En suma, el diseño constituye la mejor estrategia a seguir por el investigador para la adecuada solución del problema que tiene planteado. Desde este punto de vista podríamos considerar el diseño como «un auténtico esquema de acción» que implica, por parte del experimentador, todo un laborioso proceso de toma de decisiones. Es decir, el científico, para hacerse con la información pertinente a su hipótesis, deberá plantearse y resolver una serie de cuestiones referentes a:

- 1) La identificación de aquellos factores que de una manera u otra van a intervenir en su experimento.
- 2) La selección del tipo de medidas que va a utilizar para el registro de sus observaciones.
- 3) La correcta disposición y organización de los datos para su análisis, así como la adecuada utilización de las correspondientes pruebas de significación estadística.
- 4) La posibilidad de extender los resultados a un determinado universo o población.

Si bien cada uno de estos pasos implica una toma de decisión por parte del investigador, hemos de tener en cuenta que toda organización de un plan experimental se halla supeditada a uno de sus objetivos básicos. Este objetivo consiste, según Matalon (1969), en «evidenciar la acción de los factores experimentales, así como eliminar la influencia de los factores extraños». Sólo en base a este criterio podremos disponer de un argumento a favor de las posibles relaciones que esperamos encontrar entre las diversas variables experimentales.

Y puesto que, en última instancia, el diseño se define por su objetivo básico, hemos de concluir que toda estrategia dependerá de la propia naturaleza del problema experimental. Con ello queremos dar a entender que si bien existe en metodología experimental una serie de modelos de diseño o «diseños experimentales estandarizados», el concepto de diseño tiene, en psicología experimental, un significado mucho más amplio. Sin duda alguna, hemos de convenir que cada «experimento particular requiere la aplicación de un diseño propio». No obstante, puesto que la metodología científica ha elaborado, a lo largo de estos últimos años, una serie de «modelos de diseño básico» cuya eficacia ha sido demostrada por una gran cantidad de trabajos experimentales, nos vamos a referir, en el presente escrito, a estos modelos como pautas fundamentales para la presentación de los diseños experimentales en psicología.

Por último, cabe señalar que si bien todo diseño debe constituir un «adecuado esquema de acción», también deberemos considerarlo como una buena técnica de control del «error experimental». Desde esta perspectiva, sin tener en cuenta una serie de consideraciones relativas al costo y al esfuerzo, encontraremos entre los diversos diseños analizados importantes diferencias en cuanto a la reducción del error experimental. De ahí, que la mayor o menor precisión que podamos conseguir en la valoración de nuestros resultados depende, en definitiva, de la correcta elección del diseño.

2. EL DISEÑO DE EXPERIMENTACIÓN CLÁSICO Y SUS POSIBLES ALTERNATIVAS

El diseño simple de experimentación, o como algunos prefieren llamarlo, «Diseño clásico de una sola variable», es aquel que hemos heredado de la ciencia física y, en el que sólo se puede variar un solo factor a la vez, manteniéndose constantes el resto de los factores. Este principio de experimentación, válido para las ciencias físico-naturales, es aplicable a situaciones de constancias simples, de regularidades causa-efecto, a situaciones de un total control de condiciones, etc. Sin embargo, estas situaciones, de máximo control y rigor, no suelen darse en las ciencias de la conducta y mucho menos en psicología. Concretamente en nuestra ciencia, la mayoría de las variables se hallan directamente relacionadas con el resto, y el hecho de variar una de ellas implica, automáticamente, la variación de las restantes.

Por esta razón el principio de experimentación clásico tal como lo hemos definido antes plantea, evidentemente, serias dificultades cuando se intenta aplicarlo a la psicología. Podríamos señalar que son dos las dificultades más importantes que implica la aplicación de este esquema clásico a nuestra ciencia, y que han sido analizadas por una serie de autores con el fin de conseguir alguna solución. Vamos, a continuación, a exponer los dos problemas que resultan de la aplicación del modelo de experimentación clásico a la psicología.

En primer lugar, este modelo exige que conozcamos, con el fin de evitarlas, las posibles concomitancias existentes entre los diversos factores que inter-

vienen en el fenómeno que observamos. Como ya hemos señalado anteriormente, en las ciencias de la conducta, la variación de un factor repercute en la variación de otros muchos. De esa forma, si sólo contásemos con el esquema de experimentación clásico, podríamos llegar a conclusiones un tanto ambiguas e imprecisas en relación al efecto de las variables estudiadas. Este problema ha sido tratado por algunos autores (Fischer, 1935; Solomon, 1949), quienes han propuesto unos planes adecuados para el análisis de las posibles interacciones entre factores. En todo caso, siempre es aconsejable que en psicología se tengan en cuenta estas posibles interrelaciones entre las variables a fin de evitarlas, al menos artificialmente, con una adecuada técnica de control.

De otra parte, la consideración del diseño clásico de experimentación y su aplicación en psicología nos plantea un segundo problema, aunque éste haya preocupado menos a los interesados por el método. Me refiero a las posibles interacciones que pueden surgir, en un experimento concreto, entre los tratamientos y los fenómenos observados. Este problema no afecta en modo alguno a la propia validez interna del experimento,¹ sino más bien al nivel de generalización de los resultados. Cualquier conclusión que obtengamos de un experimento deberá interpretarse, exclusivamente, en función de las modalidades estimulativas tal como han sido aplicadas en este caso concreto. Esto nos lleva, lógicamente, a pensar en la artificialidad de los planes experimentales, y que, consecuentemente, los sujetos reaccionen más bien a dicha artificialidad que a las propias condiciones reales. En este caso es arriesgado, por no decir incorrecto, extender las conclusiones más allá de la propia situación experimental. Vamos a poner un ejemplo. Supongamos que estamos interesados en estudiar el efecto del «incentivo» sobre «el rendimiento» en un tipo de tareas relativamente fáciles. En este caso, puede ocurrir que los sujetos que ejecuten una tarea para la obtención de una recompensa muy alta reaccionen más bien a la rareza de esta situación (en condiciones de vida normal esto no suele ocurrir) que al propio valor del incentivo. Esto nos plantea, sin duda, una problemática bastante difícil de resolver y que tan sólo, hasta el momento, Brunswik (1947) y Hammond (1954) han puesto de manifiesto, pero cuya solución definitiva está aún lejos de ser alcanzada.

Frente a las desventajas y limitaciones que ofrece el diseño clásico han reaccionado, decisivamente, dos grandes teóricos, Fisher (1935) y Brunswik (1947). Ambos han puesto de relieve la incapacidad de los esquemas clásicos para áreas tan complejas e interaccionadas, y a su vez, han propuesto posibles vías de solución. Para Fisher (1935), los experimentos de la física clásica, de variación de un solo factor, servirían para familiarizar al estudiante con situaciones elementales para el conocimiento del funcionamiento de leyes simples en campos donde la incidencia de los factores restantes quedase totalmente li-

1. La validez interna se interpreta como la mínima exigencia que podemos pedir a un experimento. Es decir, suponer que la diferencia existente entre los resultados sea consecuencia de la acción de los tratamientos experimentales.

mitada. Sin embargo, estas situaciones suelen ocurrir raras veces en investigaciones de tipo biológico, en las que se desconoce, de antemano, hasta qué punto la variación de un factor afectará la variabilidad de otros que mantienen una estricta relación con el fenómeno estudiado.

Tanto Fisher como Brunswik muestran su desacuerdo con el diseño experimental clásico, y ambos preconizan su revisión. Sin embargo, la solución adoptada por estos dos autores no es, en modo alguno, coincidente.

Concretamente, Fisher (1935) propuso la técnica del «análisis de la variancia» como una posible solución al problema. Hemos de tener en cuenta que Fisher se mueve dentro del área de investigación agrícola, y por tanto, dicho procedimiento parece ser, en principio, viable. Por su parte, Brunswik y su escuela, interesados más bien por la temática perceptiva, recusan la solución propuesta por Fisher. Para Brunswik la utilización de un diseño de acuerdo con la técnica de Fisher, no convierte el diseño en un «plan menos sistemático que el diseño clásico».² Por tanto, en opinión de Brunswik, lo que Fisher ha hecho es desarrollar un diseño mucho más complejo que el clásico, ya que se tienen en cuenta la acción simultánea de varios factores, pero no ha tocado el fondo de la cuestión. Según Brunswik lo único que ha hecho Fisher ha sido sustituir una condición por varias, y por tanto, el problema no ha quedado resuelto. Por dicha razón Brunswik (1947) apunta hacia una revisión más radical de la metodología experimental.

Para Brunswik las condiciones experimentales deberían representar una muestra aleatoria del universo de condiciones al que posteriormente se van a generalizar los resultados. De esa forma se evita la artificialidad de la experimentación mediante una correcta selección de las variables. El empleo por parte de Brunswik del término «plan representativo» (como opuesto al de «plan sistemático» que atribuye a Fisher) no significa otra cosa que la trasposición del principio estadístico del muestro de los sujetos, al muestreo de las situaciones y condiciones experimentales. Por tanto, lo que en definitiva propugna Brunswik es que se aplique el muestreo, tal como viene haciéndose con los sujetos, a las diversas situaciones estimulativas. «Es evidente —añade Brunswik— que la significación estadística de un resultado debe ser estudiada en estas dos direcciones». Con este enunciado se establece que los resultados deben ser analizados, no sólo en relación a la población de sujetos, sino también, en relación a las situaciones estimulativas (e.d. en relación a su generalidad ecológica). Con ello Brunswik quiere dar a entender que el investigador no sólo debe estar interesado por la cantidad de variables que intervienen en un experimento, sino también «por las modalidades de su variación y covariación». Esta sería, en síntesis, una de las críticas más fuertes hechas por Brunswik a Fisher, quien únicamente se hallaría interesado por el número de variables que intervienen en un experimento.

2. Plan sistemático o diseño sistemático es aquél en el que se manipulan las condiciones externas de variación; es decir, las variables independientes.

De otro lado, es evidente que si no se parte de la independencia de los factores experimentales, no se pueden aplicar las técnicas de análisis elaboradas por Fisher. En cambio, en el caso de los «diseños representativos» propugnados por Brunswik se trataría más bien de calcular las correlaciones existentes entre los factores entre sí, y entre éstos y los datos experimentales. Estamos, pues, ante un problema de validez externa, de representatividad, que no sólo afecta a los sujetos experimentales, sino también, a las mismas condiciones. Estas deberán reflejar, por consiguiente en el experimento, una situación real y no la artificialidad que solemos encontrar en los trabajos experimentales.

A pesar de los poderosos argumentos esgrimidos por Brunswik, como bien señala Matalon (1969), éstos no han encontrado el suficiente eco en el campo de la experimentación psicológica. Ello puede atribuirse, quizá, a que no todos los experimentalistas se hallan igualmente interesados por la realidad inmediata, puesto que su único interés estriba en el estudio de las leyes generales y no de los casos concretos. Yo por mi parte, añadiría, que entraña una gran dificultad el llevar a cabo el programa de selección de condiciones, sobre todo para áreas psicológicas diferentes de las que Brunswik ha trabajado.

Un último argumento, que quizá justifique en parte la poca afición que muestran los experimentalistas por el uso de los planes representativos, queda materializado en la frase de Matalon al afirmar que «el poder y la elegancia de las técnicas del análisis de la variancia, la ingeniosidad que puede desplegarse para construir un plan bien adaptado a un problema, ejerce una atracción más poderosa que las técnicas de correlación».

Estas pueden ser, quizá, las principales razones que nos justifiquen el porqué la línea iniciada por Brunswik en el campo de la percepción no hayan tenido la aceptación que se podía esperar.

3. EL PROBLEMA DE LA «VALIDEZ» DE LOS DISEÑOS EXPERIMENTALES

Campbell (1957)³ distingue dos tipos principales de validez en los diseños experimentales. Por un lado habla de la validez interna, y por tal entiende al hecho de si «la estimulación experimental ha determinado claramente las diferencias constatadas en la situación concreta de tal experiencia». Tenemos, por tanto, en la validez interna la efectiva acción de los tratamientos experimentales que deberá reflejarse en la diversificación de los resultados. El propósito principal de todo investigador es, sin duda alguna, conseguir dicha validez, ya que gracias a ellas puede lograr una comprensión empírica de sus presupuestos hipotéticos. Cuando, por otra parte, habla de la validez externa de un diseño, Campbell se refiere a la representatividad o generalidad de los resultados de un experimento. Dicha validez se relaciona, por consiguiente, a las clases de pobla-

3. Se puede consultar también la obra de D. K. Campbell y J. S. Stanley, *Experimental and quasi-experimental Designs for Research*. Rand McNally, Chicago, 1963.

ciones a las que pueden generalizarse los resultados. Y estas poblaciones no deben limitarse exclusivamente a las de los sujetos, sino que incluyen las relativas a las de los tratamientos e incluso a las de las «pautas o patrones estimulativos».

Muchas veces, nos indica Campbell, cuando se pretende controlar aquellos factores extraños que comprometen la validez interna, se influye negativamente sobre la validez externa, siendo, pues, difícil conseguir un equilibrio entre ambas clases.

Campbell, al hablar de los factores extraños que pueden llegar a afectar la validez interna de un experimento, enumera los siguientes: a) la historia o circunstancias que pueden ocurrir desde que se produce la situación hasta que se registra el fenómeno; b) la maduración; c) el efecto de la medida, que puede llegar, incluso, a alterar los resultados; d) las oscilaciones que pueden tener los instrumentos de medida; e) la regresión estadística; f) el efecto de la propia selección, y g), la pérdida de sujetos durante el experimento. Cada uno de estos factores, posible fuentes de error, pueden ser fácilmente neutralizados mediante la introducción en el diseño de un grupo control. Esta ha sido una de las principales causas por las que autores tales como Solomon (1949) y Boring (1954), elaborasen planes de experiencia con grupo control. Planes, que, como veremos, podemos denominar correctamente «experimentales».

La principal dificultad que podemos encontrar en la utilización de los diseños con grupo control, es su ineficacia para el control de la posible interacción existente entre los diversos factores experimentales. La posible interacción entre las diversas variables compromete, fundamentalmente, la validez externa del diseño. En efecto, los compuestos estimulativos suelen obedecer a este patrón interactivo de tal forma que en la mayoría de los experimentos es más interesante estudiar la acción de la interacción que la de los factores independientes. La presencia de tales interacciones puede afectar notablemente los resultados a efectos de generalización, siempre y cuando dicha interacción sea consecuencia de la misma disposición de las condiciones experimentales.

El diseño experimental de cuatro grupos de Solomon (1949), así como el procedimiento del análisis de la variancia desarrollado por Fisher (1935), ofrecen al investigador una posibilidad de estudiar el posible efecto interactivo existente entre las variables. Las principales ventajas de cada uno de estos dos sistemas las analizaremos más adelante cuando presentemos, con más detalle, los diversos diseños experimentales. Sólo indicaremos ahora que la eficacia de estos diseños es tal, que pueden ser considerados como técnicas ideales de experimentación para la mayoría de los trabajos de psicología experimental; sobre todo para aquellos casos en los que estamos interesados en una máxima aproximación al contexto real de la conducta. En efecto, los diseños factoriales desarrollados por Fisher son procedimientos que en un futuro no muy lejano serán imprescindibles para cualquier trabajo experimental.

Nos queda, por último, el problema referente a la generalización de los resultados a las condiciones experimentales. Si mediante un adecuado procedi-

miento de aleatorización podemos generalizar los resultados a la población de sujetos de donde procede la muestra elegida, no podemos afirmar lo mismo en relación a las condiciones experimentales. «Si la disposición de los tratamientos experimentales —afirma Hammond (1954)—, y no solamente su número, no es concebida en la experiencia en función de las condiciones a las que se quiere generalizar los resultados, contamos con un estrecho margen para generalizar».

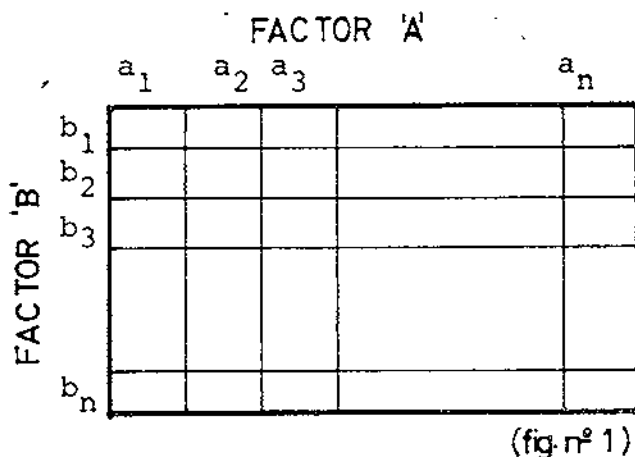
Como ya hemos indicado anteriormente, Brunswik (1947, 1956) ha propuesto el muestreo de las condiciones experimentales, de la misma forma como, normalmente, viene haciéndose con los sujetos. De esta forma se aseguraría la posibilidad de generalización de los resultados a todas las condiciones estimulativas antecedentes, y no exclusivamente a la forma cómo han sido presentados en el experimento. No obstante, hemos de señalar, que la experimentación se halla actualmente más interesada por la regularidad y la constancia, y por esta razón se parte siempre de situaciones muy precisas, ya que sólo se pretende buscar la relación existente entre éstas y el conjunto de fenómenos estudiados. En efecto, el investigador no suele estar interesado por la multiplicidad y la variedad que representaría el total de posibles condiciones que en un momento dado pudiesen afectar los resultados. El investigador prefiere limitarse a trabajar con sistemas y procedimientos que le ofrezcan una mayor libertad de manejo y una mayor eficacia en el análisis de los resultados.

4. DISEÑOS SIMPLES Y DISEÑOS FACTORIALES

Desde que Fisher (1935) estableció las bases para el desarrollo de los diseños factoriales, los diseños se suelen dividir en «simples» (o de factor único) y «factoriales» (de dos o más factores).

La razón básica de esta clasificación consiste en si en el experimento interviene una sola variable (con dos o más valores), o más de una variable. En el primer caso se trataría de un diseño de corte clásico, y con él se pretendería conocer el efecto de un solo factor sobre el fenómeno experimental. Ello implica, consecuentemente, un adecuado control del resto de posibles variables extrañas capaces de afectar nuestros resultados. Variables que como ya hemos visto han sido enumeradas por Campbell (1957).

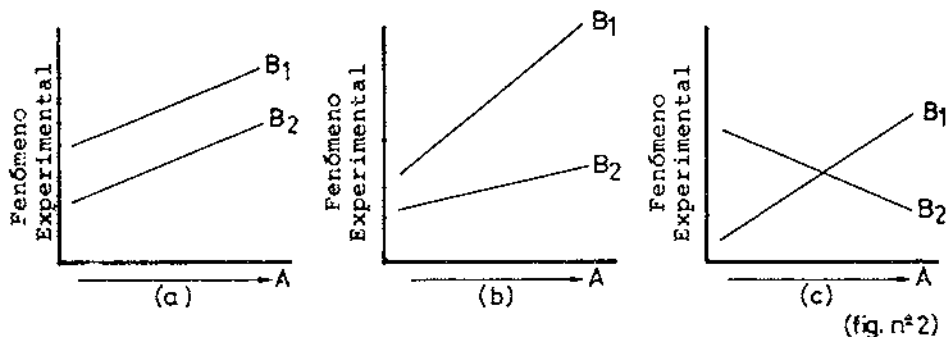
En el caso de diseños factoriales se trataría conocer hasta qué punto la variación sistemática de un factor depende o no de la forma como varían los restantes. Esto nos lleva a un tipo de plan experimental en el que aparecen todas las posibles combinaciones de cada una de las modalidades de un factor con la de los restantes. Así pues, con un diseño factorial se forman tantos grupos de sujetos, cuantas posibles combinaciones puedan establecerse entre los diversos tratamientos. Podríamos esquematizar la disposición factorial mediante la siguiente figura.



Las ventajas de esta forma de disposición son obvias. En primer lugar, cada grupo o celda representa una determinada combinación entre los diversos valores que presentan las variables. Por otra parte, la comparación de las columnas en sus resultados totales, nos permite probar la hipótesis relativa al primer factor. Lo mismo puede decirse con respecto a las filas. Así, podemos probar la acción de cada una de las variables de una forma separada (efectos principales).

Pero el diseño factorial ofrece una nueva información que no podría obtenerse con la aplicación de otro tipo de diseño. En efecto, con este diseño podemos comprobar si el efecto de un determinado valor de la variable A, varía a medida que se presenta combinado con los respectivos valores de la variable B; o si por el contrario no queda afectado por éstos. En el caso de que la variación sea efectiva, podemos esperar que existe una interacción entre ambos factores. Hemos de tener en cuenta que el concepto de interacción es muy importante y que sólo puede ser comprobado mediante el uso de diseños factoriales.

Podemos, pues, afirmar, en términos generales, que la interacción entre dos factores se produce cuando la variación del fenómeno experimental en función de uno de ellos depende de los respectivos valores que toma el otro. De forma gráfica podemos representar la interacción de la siguiente manera:



En el primer caso (figura n.º 2, a), nos encontramos con una situación en la que no existe interacción alguna entre los dos factores. Esto queda reflejado en la gráfica con dos rectas paralelas; es decir, los datos experimentales aumentan, en la misma dirección y pendiente en función del factor A, para cada valor que toma el factor B. Sólo existe una diferencia entre los valores que se obtienen en función de A y B, que se mantiene a lo largo de la relación. En este caso, puede inferirse que la influencia de ambos factores se suman sin más.

En las figuras (b) y (c), tenemos representados dos ejemplos de interacción. En el primer caso (b), nos encontramos que la acción del tratamiento A sobre los resultados es más acusada cuando se combina con el primer valor de B (B_1). En el tercer gráfico (c) tenemos una interacción inversa. El efecto de A sobre los resultados se invierten según se combine con B_1 o B_2 .

De la consideración de estos ejemplos hipotéticos podemos concluir que no tiene sentido hablar de la influencia de un solo factor sin que se tenga en cuenta las modalidades del segundo. Así pues, para un experimento que siguiera la relación establecida en (c) sería arriesgado interpretar el efecto principal de B, sin la previa consideración de A; ya que como hemos visto, la acción de B, varía según los diversos valores de A. Por tanto, con la simple consideración de los efectos principales nos veríamos en la posibilidad de obtener resultados totalmente erróneos. Todo ello nos hace caer en la cuenta de la importancia de la interacción (efectos factoriales secundarios), y de la necesidad de utilizar un adecuado procedimiento de análisis.

De lo anterior podemos concluir que es aconsejable, siempre que ello sea posible, utilizar los diseños de tipo factorial; a pesar de que su uso actual sea muy limitado en el área de las ciencias de la conducta.

Con la realización de investigaciones que impliquen diseños de factor único, nos veremos obligados a estudiar la acción de las variables de una forma aislada e independiente. Es decir, tendremos que ir probando, por separado, el efecto de cada una de las variables sin que con ello podamos llegar a conclusiones precisas de cómo, en una situación dada, actuarían si se presentasen juntas. Con el empleo sistemático de diseños simples no podemos tener la seguridad de si la acción total de los factores se puede considerar como la suma de sus efectos independientes. De ahí que los diseños de factor único posean muy poca capacidad explicativa, sobre todo para la investigación psicológica.

5. EXPERIMENTOS BIVALENTES, MULTIVALENTES Y FACTORIALES

En el presente escrito vamos a presentar los diseños en función del número y la cantidad de tratamientos. Por esa razón, hemos considerado oportuno presentar las principales formas de experimentación de acuerdo con el número de factores. Esta categorización de los experimentos en bivalentes, multivalentes y factoriales nos va a servir de marco de referencia para la descripción de cada uno de los diseños experimentales.

De otro lado, omitimos aquel tipo de experimentos denominados muy acertadamente por Campbell (1957) como «cuasi-experimentos», en los que solamente interviene un solo grupo experimental. Debido a que la mínima exigencia de rigor requiere la presencia de «un grupo control», vamos a partir de los experimentos bivalentes, en los que, como veremos, se cumple básicamente con este requisito.

5.1. Experimentos bivalentes

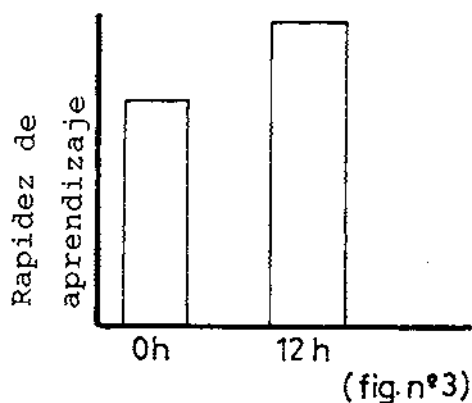
La razón por la que presentamos en primer lugar los experimentos bivalentes es básicamente de carácter histórico. En efecto, hemos de admitir que gran parte de los experimentos que se llevaron a cabo durante los primeros decenios de nuestro siglo, fueron de tipo bivalente.

En estos experimentos se suele combinar un grupo experimental con otro control, que nos sirve para neutralizar las posibles variables extrañas que pueden llegar a falsear la validez interna del mismo. Al propio tiempo, los resultados obtenidos en el grupo control nos van a servir de criterio para la verificación de la hipótesis.

Los diseños que pertenecen a esta categoría de experimentos tuvieron una finalidad clara, ya que como señala Andrews (1948), no sólo sirvieron de método de aproximación a los problemas científicos, sino también como procedimientos para la identificación de variables. Estos diseños, que podríamos calificar de tanteo, han servido para potenciar nuevas y más sistemáticas investigaciones. Y no sólo han sido utilizados como procedimientos de ataque para introducirse en áreas totalmente desconocidas, sino que además, han demostrado su utilidad sobre todo en las fases de constitución de nuestra ciencia.

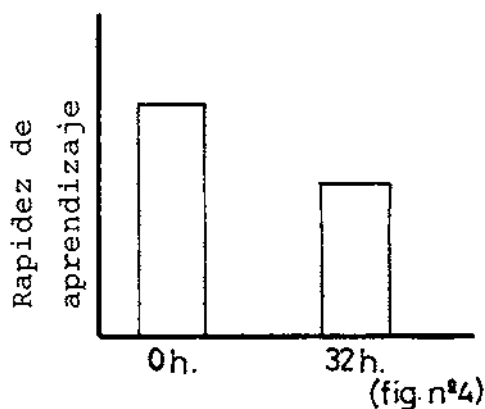
Por otra parte, hemos de señalar, en relación a los posibles resultados obtenidos mediante la utilización de tales diseños, que las posibilidades de generalización quedan enormemente restringidas, y que por tanto, las conclusiones pueden ser consideradas representativas siempre que se tengan en cuenta una serie de limitaciones básicas.

Vamos a ilustrar, mediante un ejemplo hipotético, los graves riesgos que implica el uso de este tipo de experimentos, sobre todo en el momento de interpretar los resultados. Vamos a partir de la hipótesis en la que se pretende estudiar el efecto del «impulso del hambre» (manipulado mediante horas de privación) sobre la «rapidez en el aprendizaje discriminativo». Para ello hemos elegido, de una forma arbitraria, dos valores de la variable «impulso» que podrían ser: 12 horas de privación y 0 horas de privación (grupo control). Los resultados podrían representarse mediante la figura siguiente:



Según se desprende de esta representación gráfica, es posible que concluyamos, una vez probada la significación de esta diferencia, que «a mayor privación obtendremos una mayor rapidez en el aprendizaje discriminativo». En principio dicha inferencia se apoya en datos empíricamente válidos y estadísticamente significativos.

Supongamos que, en otro laboratorio, se realice un experimento similar al nuestro, pero que en lugar de «12 horas de privación», se elijan «32 horas de privación», y se comparen los resultados, tal como lo hemos hecho nosotros, con los del grupo control (es decir, 0 horas de privación). En este segundo caso, podríamos representar los resultados de una forma similar a la nuestra (figura n.º 4):



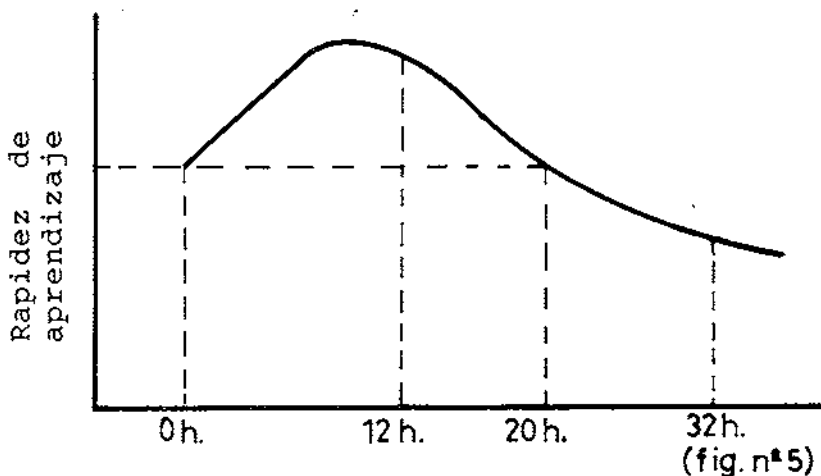
Comprobado, también, que la diferencia entre ambos grupos es significativa, se obtiene esta nueva conclusión: «a mayor cantidad de horas de privación menor rapidez en la adquisición del aprendizaje discriminativo.»

Del análisis de estos dos ejemplos hipotéticos, correctamente planificados,

llegamos a la conclusión de que los resultados referentes a una misma situación experimental son totalmente contradictorios. Analizando las posibles causas de esta palpable contraposición, podemos comprobar que ha habido por parte de las dos interpretaciones una «excesiva sobrevaloración de los resultados»; es decir, se les ha atribuido un alcance que no tenían.

Estamos, por tanto, ante un problema de validez externa, y el error que se ha cometido ha sido el de generalizar para todas las condiciones experimentales.

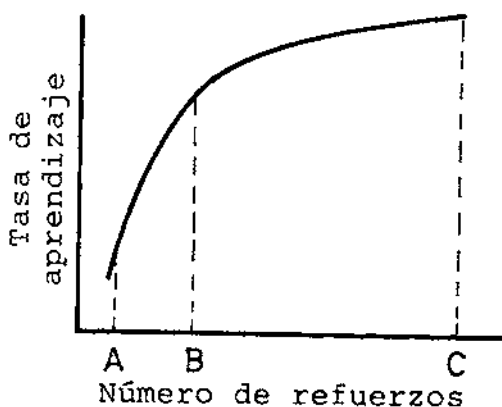
Hemos de tener en cuenta, por otra parte, que se ha partido, en ambas interpretaciones de un supuesto falso. En efecto, se ha supuesto que la relación existente entre las dos variables es lineal, sin prever la posibilidad de que dicha relación podía haber sido curvilínea, como ocurre en la figura n.º 5:



De la consideración de esta gráfica, podemos concluir que si otro investigador hubiese elegido el valor «20 horas de privación», habría concluido que no existía diferencia alguna en relación al valor «0 horas». En el caso de haber elegido un valor superior a: «32 horas», el rendimiento observado hubiese sido significativamente inferior al de «0 horas». En suma, debido a esta serie de dificultades, inherentes a los experimentos bivalentes, hemos de concluir que este tipo de experimentos poseen un carácter básicamente exploratorio, puesto que, como ya hemos indicado, muchas veces el investigador desconoce la previa relación existente entre las variables que estudia. En efecto, si las variables se relacionan entre sí según una función en «U» invertida, existen muchos valores cuyas comparaciones no van a dar diferencias significativas. Hemos de reconocer, por tanto, que los resultados positivos de experimentos bivalentes no son, ni mucho menos, completamente confirmatorios; tampoco, los resultados negativos, serán, por su parte, totalmente concluyentes.

Además, no toda relación entre variables psicológicas se resuelve en fun-

ción de una línea recta o curva, ya que se dan muchos casos en que dicha función es típicamente «exponencial»; es decir, asintótica. Este tipo de funciones las solemos encontrar en los trabajos clásicos sobre el aprendizaje, así como, en la mayoría de las investigaciones realizadas en torno a los umbrales sensitivos. Consideremos, pues, una función de este tipo, y elijamos arbitrariamente, tres valores de la misma, según viene indicado en la siguiente representación (figura n.º 6):



(fig.nº6)

Esta función representa la posible relación existente entre la variable «número o cantidad de refuerzos» y la variable «tasa de aprendizaje de una tarea simple». Supongamos que para la ejecución de un experimento concreto se eligen los valores B y C. Puesto que ambos se hallan en la asintota, no encontramos diferencia significativa alguna entre ambos. Si por otra parte, comparásemos A con B, la diferencia sería grande. De ello se deduce, que para este tipo de funciones, los riesgos de seleccionar valores no significativos son, también, muy grandes.

A título de resumen indicaremos que los peligros más importantes que presentan los experimentos bivalentes son: a) el posible desconocimiento inicial que se tiene sobre la relación existente entre las dos variables que se estudian. Según, pues, los valores que se elijan de la variable independiente, se puede llegar a resultados totalmente opuestos. b) un segundo peligro se refiere a la generalización de los resultados. Para evitar esta segunda dificultad será suficiente que se tenga en cuenta, con objeto de no falsear la generalidad de los resultados, que toda información que obtengamos en un tipo de experimento bivalente es siempre muy restringida, y por tanto, sólo aplicable a la situación concreta en la que se ha realizado el experimento. Esto nos exige una prudencia extrema en el momento de generalizar los resultados, no extendiéndolos más allá de los límites y condiciones impuestas por el marco experimental.

A continuación vamos a presentar los principales diseños que se hallan dentro del marco de los experimentos bivalentes. Todos ellos utilizan una sola

variable,⁴ y tradicionalmente, suelen utilizar un «grupo control» como base de contrastación de los efectos de la variable experimental, así como para neutralizar los posibles factores extraños que pueden llegar a contaminar los resultados.

Por otra parte, puesto que estos diseños han sido inicialmente concebidos en función de la experimentación física, los vamos a considerar como «diseños clásicos», ya que gracias a ellos se establecieron las bases para la construcción de la psicología científica.

A efectos prácticos vamos a utilizar, en la presentación de los diferentes diseños experimentales, la siguiente notación: X=variable independiente o tratamiento experimental; O=observaciones o medidas tomadas de los fenómenos experimentales; Az=asignación de los sujetos a los grupos experimentales según un procedimiento al azar; Ap= asignación de los sujetos, mediante algún sistema de apareo; E=grupo experimental; y, C=grupo control.

Vamos a pasar, pues, a la descripción de los principales diseños de «factor único», o como suelen denominarlos algunos autores «diseños clásicos». Como veremos, la mayor parte de este tipo de diseños utilizan la técnica del grupo control como sistema para lograr controlar, o al menos neutralizar, todas las posibles fuentes de variación extrañas que podrían llegar a falsear los resultados.

5.1.1. *Diseño antes-después*

En el diseño antes-después, se parte de dos grupos, uno experimental (E) y otro control (C), a los que se les asignan los sujetos mediante un procedimiento aleatorio. Gracias a este tipo de asignación se cuenta con una base razonable para pensar que los dos grupos son, teóricamente, equivalentes.

Una vez han sido formados los dos grupos, se mide alguna característica conductual relacionada directamente con el experimento. Esta medida nos puede servir de criterio para comprobar si la equivalencia de ambos grupos es efectiva, así como también, de garantía para probar si los diversos factores extraños han afectado por igual a ambos grupos. Así pues, mediante la comparación de las «observaciones antes» con las «observaciones después» podremos conocer si aquellos factores que se introducen a lo largo de la experiencia (historia, maduración, efecto de la medida, etc.) han ejercido una influencia similar en ambos grupos.

Por otra parte, puesto que el grupo experimental se halla sometido al tratamiento que se estudia, y el grupo control no, al comparar las «observaciones después» tendremos un criterio para inferir la eficacia del tratamiento experimental. Quizá uno de los aspectos más importantes tanto para éste como

4. O factor independiente. Esta es la razón por la que se les conoce, también, con el nombre de «diseños de factor único».

para otros diseños, es la simultaneidad con que ambos grupos son medidos en sus respuestas.

5.1.1.a. Esquema del diseño antes y después

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
E	Az	O_1	X_1	O_2
C	Az	O_3	X_0	O_4

La lógica formal de este diseño consiste en tomar, mediante la utilización de algún tipo de medida, los datos de rendimiento de los sujetos, antes y después de que uno de los dos grupos haya sido sometido a un tratamiento experimental (X_1), y el otro no (X_0). Las puntuaciones «antes», como ya hemos indicado, nos van a servir para comprobar la equivalencia inicial de los dos grupos con respecto a alguna característica asociada con la respuesta final. En el caso que esta equivalencia quede establecida, será suficiente comparar la medida de las observaciones O_2 y O_4 , a fin de probar nuestra hipótesis. En el caso en que la equivalencia de ambos grupos no quede demostrada será necesario realizar algún tipo de ajuste de las «medidas después», en relación a las «medidas antes», para poder, así, verificar la hipótesis (esto se consigue mediante alguna forma del análisis de covariancia).

5.1.1.b. Ventajas e inconvenientes

Gracias a la aleatorización se puede contar con una técnica capaz de controlar todas las posibles fuentes de error que afecten los resultados. El hecho de que el experimentador pueda realizar una prueba sobre la equivalencia inicial de ambos grupos le permite contar con una base razonable de que ninguna variable ha llegado a interferir la variable experimental. Así, mediante este diseño no sólo se parte de que ambos grupos son homogéneos sino que a su vez, podemos verificar la efectividad de la técnica de control.

Por otra parte, cualquier tipo de influencia externa que puedan recibir los grupos durante el transcurso de la experiencia se puede estimar mediante la comparación de las «puntuaciones antes y después». Al mismo tiempo, puesto que las observaciones «antes y después» ocurren en una misma secuencia temporal, poseemos una nueva garantía de que los resultados no han sido alterados como consecuencia de los cambios habidos en la situación.

Una de las limitaciones más importantes asociada a este tipo de diseño se deriva del hecho de la aleatorización. De una forma general podemos afirmar que todos aquellos diseños que utilizan la aleatorización como técnica de formación de grupos, requieren muestras bastante grandes de sujetos, para conseguir así un efectivo control de todas las fuentes extrañas de error.

De otro lado, el hecho de tomar medidas iniciales a los sujetos puede predisponer al grupo experimental al posible tratamiento que le podamos aplicar. Esto quizá puede comprometer su validez externa. Los resultados de tales experiencias se pueden generalizar siempre y cuando las poblaciones a las que queremos extender los resultados «hayan sido sometidas, también, a las mismas pruebas iniciales». Esta condición no suele cumplirse con frecuencia, y ésta es quizá la principal limitación que encontramos en estos diseños.

5.1.2. *Diseño de dos grupos al azar*

Basados en los planes experimentales elaborados por Fisher (1935) en el campo de la agricultura, tenemos un tipo de diseño en el que no es necesario tomar ningún tipo de medida «antes». Es suficiente, pues, para estos diseños partir de la equivalencia de los grupos, y ello se consigue mediante un procedimiento de asignación al azar. Desde el punto de vista metodológico, este tipo de diseños ha sido muy eficaz en psicología, ya que nos ofrece un procedimiento idóneo para la inferencia de la hipótesis, y a su vez, un control de todo posible efecto de las «medidas antes» sobre las «medidas después». Si bien, hemos de reconocer que mediante este diseño no podemos llegar a conocer este posible efecto, no obstante a efectos de validación de la hipótesis ofrece mucha más garantía y, consecuentemente, mayor validez interna.

5.1.2.a. *Esquema del diseño de dos grupos al azar*

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
E	Az	—	X_1	O_1
C	Az	—	X_0	O_1

Mediante el procedimiento de muestreo al azar, tenemos la garantía de que ambos grupos (el experimental y el control) son equivalentes antes de la aplicación del tratamiento. Si bien con este tipo de diseño no podemos medir el efecto que un determinado «pretest» de selección puede ejercer sobre los resultados finales, hemos de reconocer que lo que principalmente importa al experimentador es conocer la influencia efectiva del tratamiento (Payne, 1951), y que por consiguiente, toda esta serie de efectos secundarios escapan un tanto del propósito final de la investigación.

Debido a la seguridad en el control y simplicidad en la ejecución este diseño ha sido incorporado con éxito a la investigación psicológica, sobre todo, en sus fases iniciales.

5.1.2.b. *Ventajas e inconvenientes*

Partiendo de la base, gracias al empleo del muestreo al azar, de que ambos grupos siguen una distribución normal, el tipo de análisis aplicable a este diseño es relativamente sencillo. Para ello es suficiente verificar, con una adecuada prueba estadística, si la diferencia de puntajes medios existente entre los dos grupos es mayor que lo que normalmente cabría esperar si sólo hubiese actuado el azar.

De otra parte, puesto que en este diseño sólo se toman medidas de las observaciones después de la aplicación del tratamiento, se elimina el posible «error progresivo» debido al efecto de aprendizaje o fatiga, que podrían llegar a contaminar la variable experimental.

El principal inconveniente de este diseño se refiere al tamaño de la muestra, la cual deberá ser lo suficientemente grande para garantizarnos la equivalencia de los dos grupos. Asimismo, se aconseja que se investigue simultáneamente con ambos grupos, para evitar cualquier tipo de diferencia debido al tiempo o a determinadas circunstancias extrañas que pueden llegar a afectar los grupos.

5.1.3. *Diseño de grupos apareados*

El propósito fundamental de este diseño es obtener dos muestras de sujetos que, en relación a alguna actividad o característica conductual, presenten una media y una desviación estándar lo más idénticas posible. Para conseguir esto se asignan al azar a cada uno de los dos grupos pares de sujetos con características psicológicas iguales, sobre todo, las relativas a la variable de medida. De esta forma aseguramos una homogeneidad inicial entre los grupos, al tiempo que nos permite trabajar con muestras mucho más reducidas. Dada esta fuerte homogeneidad, podemos inferir con gran margen de confianza que cualquier cambio ulterior que observemos después de la introducción del tratamiento experimental será debido a éste. Existe, consiguientemente, una clara diferencia entre este procedimiento y el del azar, ya que al conseguir dos grupos más homogéneos se reduce enormemente el «error experimental», debido a las diferencias individuales (que constituyen una de las fuentes de error más importantes dentro del área de las ciencias de la conducta). Con ello se cuenta con un procedimiento mucho más sensible para detectar las diferencias mínimas entre los valores de la variable independiente.

Por último, hemos de señalar que este tipo de diseño ha sido utilizado preferentemente, en los llamados «experimentos de laboratorio», ya que con ellos se puede conseguir un mayor control de las principales fuentes de error.

5.1.3.a. Esquema del diseño de dos grupos apareados

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
E	Ap	—	X ₁	O ₁
C	Ap	—	X ₀	O ₁

En virtud del apareamiento se consigue una total correspondencia entre los sujetos que pertenecen a cada uno de los dos grupos. Ahora bien, para que el procedimiento de apareo no introduzca en el experimento ningún tipo de distorsión sistemática, se aconseja que cada uno de los sujetos que componen el par sea asignado a uno de los dos grupos mediante un sistema al azar.

A continuación vamos a presentar los principales procedimientos de apareo que podemos utilizar en la práctica experimental.

5.1.3.b. Apareo mediante variable de correlación

Mediante esta técnica se forman dos grupos con características similares en relación a una variable directamente vinculada con la variable de medida (variable dependiente). Cuanto mayor sea la correlación existente entre ambas variables, se conseguirá una mayor equivalencia entre ambos grupos.

El principal problema que se plantea al utilizar esta técnica se refiere a la selección de la variable de apareamiento. En efecto, de ello depende el éxito o no del procedimiento. El criterio consiste, como hemos apuntado antes, en elegir un aspecto de la conducta directamente relacionado con la variable dependiente. En el caso que estudiásemos, por ejemplo, el efecto de la práctica sobre el rendimiento motriz, se conseguirían formar dos grupos equivalentes pasando, previamente, a los sujetos un test de coordinación senso-motora. A continuación en base a los puntajes obtenidos por los sujetos en este test formaríamos dos grupos con una media y una desviación estándar igual. En efecto, en este caso, se parte del supuesto que dicho test se halla altamente correlacionado con el rendimiento motriz.

En otros casos, la equivalencia entre grupos se suele conseguir mediante la utilización de variables tales como el sexo, la edad, educación, nivel socio-económico, etc. Todos estos criterios son buenos siempre que como ya hemos indicado, se pruebe la existencia de una relación estrecha con el aspecto conductual que se desea estudiar.

5.1.3.c. Apareo por pares

Podemos, también, formar dos grupos homogéneos asignando al azar a cada uno de ellos pares de sujetos con características psicológicas idénticas. Si por ejemplo, deseamos formar dos grupos equivalentes en relación a la ca-

pacidad intelectual, podríamos seleccionar pares de sujetos con un mismo C.I., y los asignaríamos, a continuación, al azar a cada uno de los dos grupos. Si además los apareásemos teniendo en cuenta la edad, sexo, rendimiento escolar, etcétera, tendríamos una mayor garantía sobre la homogeneidad de ambos grupos, pero cada vez quedaría más restringida la posibilidad de encontrar pares de sujetos con las mismas características psicológicas.

Otro aspecto a tener en cuenta en relación a esta técnica, es que con su aplicación se obtienen grupos de igual tamaño. Esto es muy interesante para la posterior aplicación de pruebas estadísticas.

5.1.3.d. *Apareo según un rendimiento previo*

En muchos trabajos se ha utilizado el sistema de aprovechar los primeros ensayos como criterio de apareamiento. Así pues, a partir de los puntajes obtenidos en el primer o primeros ensayos, se forman dos grupos de sujetos homogéneos. Mediante esta técnica podemos, incluso, eliminar aquellos sujetos que se aparten mucho de la norma del grupo.

Otra de las ventajas que obtenemos mediante la utilización de este procedimiento es que los sujetos adquieren cierta práctica con el tipo de tarea que va a ser objeto de estudio. Con ello se elimina un posible factor de novedad que podría llegar a interferir los resultados.

Una posible alternativa a este procedimiento consistiría, como suele hacerse en algunos experimentos sobre aprendizaje, en eliminar, durante esta fase de entrenamiento, aquellos sujetos que no hubiesen alcanzado un determinado nivel de eficiencia. Si establecemos, por ejemplo, que los sujetos en esta fase previa deben completar un trabajo en 30 segundos (con una variación de más o menos 5 segundos), podemos ir seleccionando sólo aquellos que han alcanzado dicho criterio. Luego los podemos asignar al azar a uno de los dos grupos. En el caso de que no alcance este nivel mínimo exigido, sería rechazado para el experimento.

5.1.3.e. *Apareo mediante control acoplado*

Durante estos últimos años, se está utilizando en los laboratorios de psicología experimental, un nuevo tipo de apareo que recibe el nombre de «control acoplado». El origen de esta técnica se remonta a una serie de experimentos en los que se comprobó que si los animales hambrientos o sedientos recibían una descarga eléctrica antes de alcanzar la comida adquirirían muchas más úlceras de estómago que los que no la recibían. Esto planteó el problema sobre el origen de la úlcera de estómago. Es decir, los investigadores se preguntaron si la úlcera la provocaba el choque eléctrico, o bien la tensión creada en el animal por la situación conflictiva (tender hacia la comida y al mismo tiempo, evitar la descarga). Para resolver este problema fue necesario emplear un

animal control que recibiera la misma magnitud y secuencia temporal de la descarga, pero sin pasar por la situación conflictiva.

Brady (1958), intentó solucionar este problema mediante la técnica del «sujeto control acoplado». Para ello entrenó a unos monos para que presionaran una palanca a fin de evitar la descarga que se producía en los pies con una frecuencia de 20 segundos. Esta descarga podía ser evitada si el mono oprimía la palanca en el intervalo de los 20 segundos. Como es obvio, los monos aprendieron rápidamente esta conducta evitativa, y sólo en el caso que dejasen de presionar la barra, recibían la descarga eléctrica.

Para determinar si la úlcera se originaba como consecuencia del choque eléctrico o de la situación conflictiva, se acopló al primero un segundo mono, dentro de la misma situación. La única diferencia consistía en que este segundo animal recibía las mismas descargas, con igual intensidad y regularidad que su compañero, pero sólo éste último podía evitar la aparición del choque. Es decir, los dos recibían la misma descarga, con la excepción de que el sujeto experimental tenía el control de la situación (y por tanto estaba sometido a la tensión de evitar el choque). Esta técnica delimitó claramente el problema, ya que el sujeto control sólo estaba sometido a la descarga, y no a la tensión.

Esta técnica, a pesar de tener una serie de inconvenientes graves, tiende a ser utilizada cada vez más y, actualmente, la encontramos aplicada al estudio del condicionamiento parpebral, del refuerzo secundario, etc. (Gormezano, Moore y Deaux, 1962). A pesar de que, como acabamos de afirmar, esta técnica se está aplicando con notable éxito, no por ello ha dejado de tener fuertes detractores.

5.1.3.f. *Ventajas e inconvenientes de los diseños de dos grupos apareados*

Una de las ventajas más importantes de este tipo de diseños, es que con ellos se consigue una mayor homogeneidad de los grupos. Por otra parte, si para los diseños de grupos al azar se necesitaba un gran número de sujetos, en los diseños de grupos apareados se pueden obtener conclusiones válidas con un total de 10 a 20 sujetos.

Al mismo tiempo, puesto que existe una mayor homogeneidad, se reduce, enormemente, la variancia debida al error, y consecuentemente, se aumenta la posibilidad de detectar menores diferencias entre los valores de la variable independiente.

Uno de los inconvenientes más graves se presenta, en el caso de que se desee aparear sujetos en más de una característica. En efecto es muy difícil encontrar sujetos con una serie de rasgos similares; por tanto, queda enormemente reducida la posibilidad de formar grupos iguales. También en el momento de inferir conclusiones y generalizar resultados, nos encontramos con una serie de limitaciones.

Todo esto nos hace ser muy prudentes en el momento que decidamos utilizar este tipo de diseños.

5.1.4. *Diseño de Solomon de cuatro grupos (Solomon, 1949)*

Este diseño viene a ser, en pocas palabras, una combinación de los dos tipos de diseños que acabamos de presentar. Es decir, una combinación del diseño antes y después y el diseño de dos grupos al azar.

Este diseño nos permite comprobar, de una forma mucho más exhaustiva, no sólo al efecto del tratamiento experimental sobre la variable dependiente, sino, también, la posible interacción existente entre las «medidas antes» y el «tratamiento». Es interesante, sobre todo en experimentos de campo, poder conocer si la previa medida que hemos tomado de los sujetos, mediante la aplicación de algún test o cuestionario, interactúa con el posible efecto de la variable independiente.

5.1.4.a. *Esquema del diseño de Solomon*

<i>Grupo</i>	<i>Asignación</i>	<i>Observ. antes</i>	<i>Tratamientos</i>	<i>Observ. después</i>
E_1	Az	O_1	X_1	O_2
C_1	Az	O_3	X_0	O_4
E_2	Az	—	X_1	O_5
C_2	Az	—	X_0	O_6

Con este diseño el efecto del tratamiento (X_1) se puede probar comparando: O_2 con O_1 ; O_2 con O_4 ; O_5 con O_3 ; y, O_6 con O_3 . Si todas estas comparaciones concuerdan entre sí, la posibilidad de inferir que la acción del tratamiento ha

Tabla de datos 2x2

Tratamientos

	X_0	X_1
Preprueba	O_4	O_2
Sin preprueba	O_6	O_5

FIG. 7

sido efectiva, crece enormemente. Tenemos, pues en un solo diseño la posibilidad de obtener cuatro pruebas independientes para verificar la misma hipótesis.

Al mismo tiempo, con un adecuado análisis estadístico, este diseño nos permite probar el efecto del tratamiento sin la presencia de la «pre-prueba»; el efecto del tratamiento con la presencia de la «pre-prueba», y por último, la posible interacción entre la «pre-prueba» (o medidas antes) y el «tratamiento».

Por otra parte, mediante una adecuada disposición de los resultados, podemos convertir este diseño en un plan factorial. En este caso las variables independientes serían: 1) la pre-prueba; 2) el tratamiento. Llevando, pues, los resultados a una tabla de doble entrada, podremos analizar y considerar los datos desde una nueva perspectiva.

Mediante esta disposición, el diseño de cuatro grupos de Solomon se convierte en un diseño « 2×2 » factorial. Con el análisis de la variancia podremos comprobar, por tanto, no sólo el efecto del tratamiento, sino también, el posible efecto de la prueba inicial. Si por otra parte, comparamos las medias de resultados en diagonal, tendremos la posibilidad de comprobar la interacción entre ambas variables.

5.1.4.b. *Ventajas e inconvenientes*

El hecho de que en este diseño se utilice, como en anteriores, el procedimiento del azar para la formación de los grupos, nos ofrece la posibilidad de controlar todas las fuentes de error. Ahora bien, debido a la especial estructura del diseño, se logra controlar de una forma más efectiva la posible influencia de «las observaciones antes» sobre los resultados finales. Y no sólo se logra controlar de una forma efectiva dicha variable, sino que a su vez, podemos calcular la magnitud de su influencia.

Por todo ello podemos concluir que el diseño de Solomon constituye un procedimiento idóneo para el control del «error progresivo» que suele darse en aquellas situaciones en las que se toman medidas previas a los individuos. Por dicha razón Solomon y Lesaac (1968), lo consideran como un método ideal para el estudio de los procesos de desarrollo, en los que suelen intervenir variables madurativas y ambientales.

Por último, puesto que se trata de un diseño de cuatro grupos, necesitaremos de una mayor cantidad de sujetos. Ello nos obligará a tomar las debidas precauciones para asegurar la equivalencia de todos los grupos, en el sentido de que no exista ningún otro tipo de variable capaz de afectar sistemáticamente los resultados.

5.1.5. *Diseños de tratamientos contrabalanceados*

Como acertadamente observa Underwood (1966), el diseño de tratamientos contrabalanceados es un diseño intrasujeto. Con ello se quiere dar a entender que un diseño de este tipo implica que cada sujeto se someta a todas las

condiciones experimentales. Si bien la utilización de un mismo sujeto para todas las condiciones experimentales tiene grandes ventajas en el sentido que se reducen al máximo los errores experimentales, no obstante, suelen aparecer los «errores progresivos» que pueden llegar a falsear los resultados. Para neutralizar el error progresivo, se suelen utilizar, corrientemente, las siguientes técnicas: el contrabalanceo intrasujeto y el contrabalanceo intersujeto.

En el primer tipo de contrabalanceo se logra la neutralización del error progresivo mediante el «contrabalanceo de las condiciones» para cada sujeto. Es decir, si en una primera fase se aplican los tratamientos en un orden dado (X_1 y X_2), a continuación se aplican los mismos tratamientos pero en orden inverso (X_2 y X_1). De esta forma neutralizamos el posible efecto del error progresivo.

Esta inversión de condiciones se puede llevar a cabo o bien en el mismo sujeto, o bien, formando dos grupos diferentes de sujetos. En el primer caso tendríamos un diseño contrabalanceado intra-sujeto, y en el segundo, inter-sujeto.

5.1.5.a. Esquema de diseño contrabalanceado intra-sujeto

Tratamientos	X_1	X_2	X_2	X_1
Observaciones	O_1	O_2	O_3	O_4

Esta secuencia (X_1, X_2, X_2, X_1) se aplica a cada uno de los sujetos que constituye el grupo en estudio.

Si analizamos este diseño podemos constatar cómo la disposición de los tratamientos X_1, X_2 y X_2, X_1 obedece al propósito de equilibrar las condiciones contrabalanceándolas. De esta forma, los posibles efectos facilitadores o distorsionadores del tratamiento X_1 sobre el tratamiento X_2 durante la primera fase de la secuencia quedan contrarrestados al cambiar la disposición de los mismos durante la segunda fase.

Para un adecuado análisis de resultados, sería suficiente comparar los promedios de las observaciones O_2 y O_3 , con el O_1 y O_4 .

5.1.5.b. Esquema de diseño contrabalanceado inter-sujeto

Grupo	Tratamiento	Observ. después	Tratamiento	Observ después
I	X_1	O_1	X_2	O_2
II	X_2	O_3	X_1	O_4

Este diseño a diferencia del anterior, aplica las secuencias contrabalanceadas en grupos diferentes de sujetos. Pueden darse, en efecto, efectos de práctica (error progresivo), ahora bien, en este caso la neutralización de tales efectos se consigue mediante la formación de dos grupos. Es decir, un primer grupo recibe los tratamientos en un orden determinado (X_1, X_2), y el segundo, a la inversa (X_2, X_1). De esa forma se logra equilibrar todas las posibles influencias del tratamiento anterior sobre el posterior.

Los objetivos de este diseño son los mismos que el anterior. No obstante, hemos de tener en cuenta que para este tipo de diseño necesitaremos un número mayor de sujetos.

5.1.5.c. *Ventajas e inconvenientes*

Los diseños contrabalanceados, especialmente utilizados en el área del aprendizaje verbal, tiene, actualmente, una más amplia difusión sobre todo en los estudios referentes al aprendizaje libre y a las investigaciones sobre la memoria a corto plazo.

Una de las ventajas generales de estos diseños es que con ellos podemos utilizar un número reducido de sujetos. Ahora bien, hemos de tener en cuenta que estos diseños son apropiados sólo para aquellos casos en que el efecto de los tratamientos tiene una duración limitada, ya que de lo contrario podrían falsearse los resultados.

Por otra parte, hemos de destacar que con los diseños «intrasujetos» queda reducido enormemente el error experimental. Lo cual constituye una notoria ventaja en la práctica de investigación.

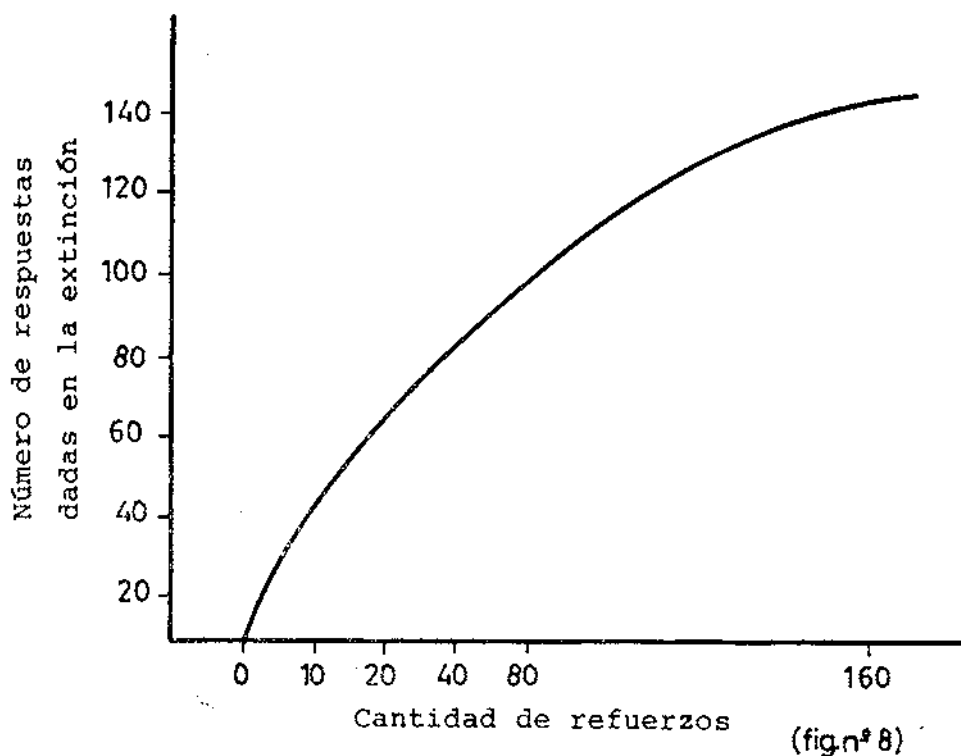
Por último, hemos de señalar, que uno de los presupuestos básicos para la aplicación de estos diseños es que las variables secundarias extrañas asociadas con el tiempo han de presentar una relación lineal. Si el efecto de tales variables sobre la conducta no es lineal, cualquier modificación o cambio que se opere de una prueba a otra, podría interferir los efectos de la variable experimental. Para estos casos, la utilización de tales diseños no sería adecuada.

5.2. *Experimentos multivalentes*

Los procedimientos de análisis, perfeccionados sobre todo por Fisher (1935), han permitido el empleo de planes experimentales de una sola variable pero con más de dos valores. Con ello podemos obtener una información más exacta sobre la posible relación existente entre las variables. Esta es la razón básica por lo que a tales experimentos se les conozca, también, con el nombre de «experimentos funcionales».

En el gráfico siguiente (figura 8), se representa un experimento multivalente, estableciéndose una relación funcional entre las variables «resistencia a

la extinción y «cantidad de refuerzos». En este experimento se han utilizado seis valores de la variable independiente, a saber: 0, 10, 20, 40, 80 y 160 refuerzos.



Los valores de la variable independiente son totalmente arbitrarios, y consecuentemente, se podía haber realizado otro experimento tomando los siguientes valores: 0, 15, 30, 120 y 240. El científico puede elegir éstos u otros valores (5), ya que, frecuentemente, él se halla interesado en la posible línea que unirá estos valores y que representará la posible función existente entre ambas variables.

Hemos de señalar que este sistema permite interpolar valores que no han sido probados experimentalmente con lo que podremos predecir, para un determinado valor de la variable independiente, cuál será la tasa de respuestas dadas por los individuos.

Estos experimentos tienen una gran importancia en psicología, ya que no sólo nos permiten establecer, con cierta precisión, la posible relación existente

5. Lógicamente aquí se podría plantear el problema del muestreo de las condiciones experimentales, a fin de garantizar la representatividad de los tratamientos. Este problema ha sido tratado ya antes y, por tanto, no vamos a insistir en el mismo.

entre dos variables, ofreciendo la posibilidad de inferir el posible efecto de un valor no probado, sino que a su vez, podemos establecer relaciones cuantitativas entre variables.

Dentro de la clase de experimentos multivalentes podemos incluir los siguientes diseños: Diseños de grupos totalmente aleatorizados, Diseños de bloques de grupos al azar, Diseños cuadrados, y Diseños de medidas repetidas (o diseños intra-sujeto con tres o más tratamientos). Hemos de reseñar que los diseños mencionados no sólo se emplean en experimentos «multivalentes de factor único», sino que además son también apropiados para experimentos de tipo «factorial». La única diferencia consiste en que las posibles variables que estos diseños controlan sean tenidas en cuenta o no en el análisis de los resultados. En otras palabras, si un diseño Latino Cuadrado nos permite controlar dos fuentes de variación, en el caso que éste se utilice en un experimento factorial serán consideradas como variables independientes.

Por último hemos de añadir que el criterio utilizado al presentar los diseños en la clasificación antes referida consiste en que cada uno de ellos utiliza un determinado procedimiento para la reducción del «error experimental». En la medida que, mediante el diseño, se logra un control más efectivo de las fuentes de variación, menor será la magnitud del error y mayor la posibilidad de probar la hipótesis (siempre, claro está, que sea válida). Desde este punto de vista el diseño puede considerarse como una buena técnica de control y como tal puede ser correctamente empleada.

A continuación vamos a describir brevemente la estructura de estos diseños, que como ya hemos indicado pueden aplicarse no sólo en experimentos funcionales sino también factoriales. Todo depende, pues, del uso que quiera hacerse de ellos.

5.2.1. *Diseño de grupos totalmente aleatorizados*

Frecuentemente se habla de los diseños en función del procedimiento que se utiliza para conseguir grupos o muestras de sujetos lo más homogéneas posible. En efecto, sólo sobre la base de la equivalencia de los grupos se pueden aplicar correctamente las pruebas estadísticas a fin de probar la hipótesis experimental. Por esa razón los experimentadores han utilizado diversos procedimientos, siendo uno de los más frecuentes la «aleatorización».

La aleatorización, si ha sido llevada a cabo de una forma efectiva, nos ofrece una garantía de la total neutralización de cualquier tipo de variable que puede llegar a afectar sistemáticamente los resultados. Con ello poseemos una base razonable para concluir que cualquier variación que no sea efecto del azar, deberá ser atribuida a la presencia de la variable independiente.

Por otra parte, este tipo de diseño implica la formación de tantos grupos experimentales cuantos sean los tratamientos experimentales que se deseen probar. A su vez, deberán asignarse a cada uno de estos grupos los correspon-

dientes sujetos de una forma totalmente aleatoria. Así, el hecho de que un sujeto pertenezca a un determinado grupo dependerá exclusivamente del azar.

5.2.1.a. Esquema de un diseño de grupos totalmente al azar de cinco tratamientos

Grupo	Asignación	Observ. antes	Tratamiento	Observ. después
I	Az	—	X_0	O_1
II	Az	—	X_1	O_2
III	Az	—	X_2	O_3
IV	Az	—	X_3	O_4
V	Az	—	X_4	O_5

Los experimentos multivalentes sirven para comparar diversos valores de la variable independiente, lo cual nos permite comprobar la existencia o no de una relación entre aquella y el fenómeno experimental. También, gracias a estos experimentos podemos llegar a establecer el tipo de relación existente entre ambas variables. Debido a esta propiedad, estos experimentos suelen llamarse «paramétricos», ya que con ellos no sólo podemos establecer el tipo de función, sino también, el sentido que toma dicha función.

Todo ello nos va a permitir un mejor conocimiento de las relaciones existentes entre los fenómenos psicológicos, y la posibilidad de una expresión matemática de dicha relación.

5.2.1.b. Ventajas e inconvenientes

Mediante la aleatorización se consigue equilibrar todas las posibles influencias sistemáticas extrañas tanto en un sentido como en otro. De ahí, que podamos esperar que los datos que obtengamos de los sujetos se distribuyan según la «ley normal», lo cual constituye una de las condiciones básicas para la aplicación de las pruebas de significación estadística.

Tampoco es necesario, para este tipo de diseño, que el número de sujetos necesarios para la formación de los grupos sea exactamente el mismo, siempre, claro está, que sea lo suficientemente grande como para que el azar pueda actuar plenamente.

Como acabamos de ver, este diseño se apoya en la técnica de la asignación al azar de los sujetos a los diferentes grupos experimentales. Si bien esto constituye una correcta forma de control, no deja de tener sus inconvenientes. Y estas dificultades no se derivan del hecho de la aplicación del azar, sino porque la utilización del azar produce una variancia inter-individual superior a la de cualquier técnica. A su vez, como ya hemos indicado anteriormente, la variancia inter-individual constituye uno de los componentes más importantes del

error experimental. Esto nos lleva a considerar que antes de su aplicación pondremos adecuadamente su uso. La razón es sencilla, ya que a mayor error experimental, menor es la posibilidad de probar la eficacia de tratamientos con diferencias mínimas. En efecto, podemos concluir que este diseño es muy poco sensible, sobre todo si las diferencias entre los tratamientos son mínimas.

Si, por otra parte, queremos asegurar una aleatorización completa, podemos asignar también al azar los tratamientos a cada uno de los grupos. Con ello aumentamos nuestra seguridad de que no ha existido ningún otro criterio, a excepción del azar, por el que un sujeto haya recibido un determinado tratamiento.

5.2.2. *Diseño de bloques de grupos al azar*

Debido a la gran variación que se opera en diseños totalmente aleatorios, los matemáticos estadísticos (Fisher, 1935), han elaborado nuevas técnicas con el fin de reducir al máximo dicha variación. Una de estas técnicas recibe el nombre de «bloques».

La primera aplicación de «bloqueo» como técnica de control y reducción de error, fue en agricultura, de donde posteriormente, se derivó el nombre. Bloque significa en agricultura un conjunto de parcelas de terreno que presentan unas características similares (fertilidad, humedad, climatología, etc.). De esta forma, todas las parcelas que se encuentran dentro de una misma área de terreno (o bloque), presentarán unos rasgos comunes, y de ahí que se consideren homogéneas en relación a los tratamientos. En efecto, dentro de cada bloque las unidades de terreno serán más idénticas entre sí que en relación a las pertenecientes a otros bloques.

Mediante este procedimiento, las diversas unidades experimentales se pueden considerar equivalentes y, por tanto, de escasa variabilidad. Como es natural, debido a la gran homogeneidad interna del bloque, podremos detectar diferencias más pequeñas entre los tratamientos.

El principio del bloqueo puede aplicarse, como es lógico, a las ciencias biológicas, así como a las de la conducta. Siguiendo esta técnica, podremos formar bloques de sujetos que presenten características psicológicas similares. Con ello habremos reducido enormemente la variabilidad dentro del bloque, y conseguiremos una mayor eficacia al probar los tratamientos.

Vamos a describir, a continuación, mediante un ejemplo, la técnica del bloqueo. Supongamos que nos interesa estudiar la eficacia de tres métodos de enseñanza, A, B y C, mediante la utilización de la técnica del bloqueo. Para ello formaremos dos bloques de sujetos, de acuerdo con unas características psicológicas directamente relacionadas con el fenómeno experimental (en este caso, el rendimiento escolar). Así, incluimos en un primer bloque todos aquellos sujetos con un C.I. de 110, y en un segundo bloque a sujetos con un C.I. de 90. Como es lógico, dentro de cada bloque la homogeneidad de los sujetos con respecto a la capacidad intelectual es completa.

Descripción gráfica del sistema (figura n.º 9):

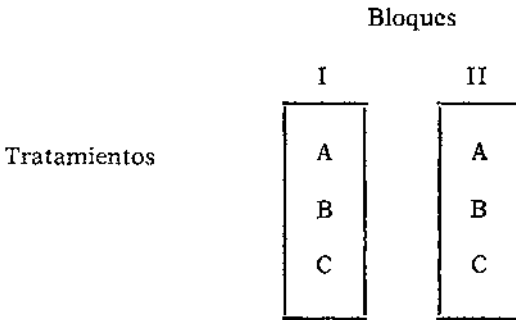


FIG. 9

Una vez formados los bloques, de acuerdo con un rasgo psicológico, dentro de cada bloque formaremos tantos grupos como tratamientos. Para la formación de tales grupos utilizaremos una técnica de asignación al azar. Con ello aseguramos el control de otras posibles fuentes de error existentes dentro del bloque.

Gracias al procedimiento del bloqueo, separamos de la variación debida al error una de sus fuentes: la debida al sistema de bloqueo. De esta forma, contamos en psicología con un procedimiento que nos va a permitir estudiar con más eficacia la acción de los tratamientos.

5.2.2.a. Esquema del diseño de bloques de grupos al azar

Bloque	Grupo	Observ. antes	Tratamiento	Observ. después
I	1	—	X_a	O_1
I	2	—	X_b	O_2
I	3	—	X_c	O_3
II	1	—	X_a	O_4
II	2	—	X_b	O_5
II	3	—	X_c	O_6

Esta forma de diseño tiene mucha utilidad en psicología debido a que las observaciones conductuales varían mucho de un sujeto a otro. Si logramos, pues, controlar mediante el sistema de bloqueo alguna de estas fuentes de variación inter-individual, lograremos grupos de sujetos mucho más homogéneos, y a su vez, una considerable reducción del error experimental.

El razonamiento lógico en que se apoya dicho procedimiento consiste en eliminar, mediante el bloque, una fuente o más de variación. De esta manera aumentamos la posibilidad de poder probar la hipótesis experimental.

5.2.3. El doble bloqueo

Con la técnica del bloqueo, como hemos visto, se reduce sensiblemente el error experimental al agrupar los sujetos de acuerdo con alguna característica psicológica directamente vinculada con la variable de respuesta. Ahora bien, siguiendo con este mismo criterio podemos avanzar un paso más al intentar la utilización simultánea de dos sistemas de bloques. En efecto, con ello seríamos capaces de controlar dos fuentes de error al mismo tiempo.

En principio podríamos utilizar dos sistemas de bloqueos de diferentes dimensiones. Ahora bien, dado que esto supondría el empleo de una serie de análisis estadísticos complejos, en la práctica, se ha optado por el empleo de «sistemas de bloques que se correspondan entre sí y que, a su vez, guardan simetría con la cantidad de tratamientos». En efecto, si hemos decidido utilizar un sistema de tres bloques para una determinada característica, y también un segundo sistema con otros tres bloques, forzosamente deberemos aplicar tres tratamientos distintos. Con ellos se logra formar «matrices cuadradas», es decir, con una misma cantidad de bloques, por sistema, que tratamientos se empleen en el experimento.

Dada esta particular disposición, los diseños que utilizan la técnica del «doble bloqueo» suelen recibir el nombre de «Cuadrados latinos», nombre derivado de un antiguo pasatiempo de carácter matemático.

Un diseño Cuadrado latino está formado por dos sistemas de bloques de «n» categorías o dimensiones cada una, de tal manera que cada bloque de un sistema contiene una unidad experimental (o grupo) perteneciente a cada uno de los bloques del otro. Estos dos sistemas son conocidos, usualmente, con el nombre de «filas» y «columnas». En efecto, cada «unidad» o «grupo experimental» se encuentra una y sólo una vez en cada fila y en cada columna. Dada esta especial disposición (en matriz cuadrada) del diseño, sólo podremos utilizar, como ya hemos señalado antes, un número de tratamientos igual al número de categorías o bloques por sistema. Vamos a ilustrar dicha técnica mediante un ejemplo.

Siguiendo con el caso anterior, podíamos haber decidido establecer un sistema de tres bloques para la característica «aptitud intelectual». Así, se podían haber formado tres bloques siguiendo este criterio: a) un primer bloque para sujetos con un C.I. de 85; b) un segundo bloque para sujetos con un C.I. de 100; y, por último, c) un tercer bloque con sujetos de 115 de C.I. Por otro lado, hemos decidido agrupar los sujetos de acuerdo con una segunda característica psicológica, que podría ser, la «tasa de retención». Para ello elegimos los siguientes percentiles: 25, 50 y 75, seleccionando, a continuación, aquellos sujetos que cumplan con estas condiciones. De esta forma, disponemos de dos sistemas de bloques, de manera que cada sujeto (o grupo de sujetos), sólo aparecerá una y sólo una vez en cada fila y en cada columna.

Dispuestos los sujetos según este criterio, un último paso consistirá

en asignar al azar, dentro de cada fila y cada columna, los tratamientos experimentales (en nuestro ejemplo, los «métodos de enseñanza A, B, y C»).

La disposición concreta del diseño cuadrado latino para nuestro ejemplo quedaría representada de la siguiente forma (figura n.º 10):

Diseño cuadrado latino

Cocientes intelectuales

		85	100	115
Retención	25	A	B	C
	50	C	A	B
	75	B	C	A

FIG. 10

Cada tratamiento (representado por las letras A, B y C) aparece una y sólo una vez, en cada fila y en cada columna. Como ya hemos indicado, la aplicación de los tratamientos a los grupos que forman un bloque es totalmente al azar. De esta forma, no existe, por parte de los sujetos pertenecientes a un mismo bloque ninguna disposición previa para recibir un tratamiento determinado. El azar, constituye de nuevo, el criterio básico de confrontación.

Hemos presentado un ejemplo, muy frecuente en psicología, en el que la pertenencia de un individuo a un grupo viene determinada por la posesión de unas características aptitudinales dadas. No obstante, se podría pensar en otro tipo de diseño de cuadrado latino en el que la asignación de los sujetos a los grupos dependiera totalmente del azar. Esto podría hacerse si las variables que constituyesen los sistemas de bloques fueran externas al sujeto (es decir, situacionales), y no orgánicas como ocurría con el ejemplo analizado antes. De esta forma se podría estudiar, por ejemplo, los rendimientos en tareas motoras, controlando por una parte el «posible efecto de experimentador» y por otra, «los momentos de ejecución de la prueba». En este caso, puesto que no se trata de dimensiones psicológicas subjetivas, podíamos haber asignado los sujetos a cada uno de los grupos, según un criterio al azar.

Podemos concluir, por tanto, que los diseños de cuadrado latino, constituyen una técnica ideal para controlar una serie de factores (gracias a los sistemas de bloques), con lo que conseguimos una notable reducción del error experimental. Esto se debe a que la variación producida por la división en bloques no afecta la variancia del error.

5.2.3.a. *Ventajas e inconvenientes del sistema de bloques*

Las ventajas que nos ofrece la técnica de bloques, ya sea simple o doble, son sin duda de gran importancia. Con esta técnica logramos homogeneizar grupos de sujetos en relación a una o dos categorías, lo cual supone una enorme reducción de la variancia del error, puesto que en el análisis de los resultados la variación debida al sistema de bloques queda aislada de la variación del error. El error experimental se calcula, en este tipo de diseños, a partir de la variación de los sujetos dentro de cada bloque. Así logramos una mayor precisión en el momento de comparar los efectos de los tratamientos.

Una de las dificultades principales que nos plantean estos diseños es que los grupos han de formarse con una cantidad igual de sujetos. Esto es muy complejo, sobre todo si se trata de controlar variables orgánicas. A pesar de todo, su uso es muy conveniente en psicología, ya que constituye uno de los procedimientos de experimentación más seguros, debido a la posibilidad de control que nos permite.

Hemos de añadir una última aclaración. Si bien, como hemos indicado, mediante este tipo de diseños reduciremos sensiblemente el error experimental, ello no quiere decir que lo eliminamos totalmente; en efecto, por más que homogeneicemos los sujetos, no puede esperarse que dentro de la situación experimental todos vayan a reaccionar de una misma forma, y, por tanto, aparecerán irremediablemente las diferencias individuales.

5.2.4. *Diseño de medidas-repetidas o intra-sujeto*

Como ya hemos indicado, las diferencias individuales constituyen una de las principales fuentes de error. Las técnicas estudiadas hasta ahora tenían como uno de los objetivos básicos, el conseguir una máxima reducción del error.

Uno de los objetivos fundamentales en este tipo de diseños, consiste en separar la variación debida a las diferencias individuales de la variación total. Para lograr este objetivo, cada sujeto es sometido a todos los tratamientos que forman el experimento; así «el bloque es sustituido por el sujeto, el cual se convierte en su propio control». De esta manera, los efectos de los tratamientos son valorados al compararlos con el promedio de respuestas dadas por los individuos a cada uno de los tratamientos. Se ha eliminado, por tanto, de dicho promedio el posible efecto debido a las diferencias individuales.

Por otra parte, puesto que hemos utilizado el mismo sujeto para cada una de las condiciones experimentales, deberemos tomar un número de observaciones, por sujeto, igual al de los tratamientos. Esta es la razón por la que este diseño recibe el nombre de «medidas repetidas».

5.2.4.a. Esquema del diseño de medidas repetidas

Sujetos	Tratamientos				Observaciones			
1	X ₁	X ₂	X ₃	X ₄	O ₁	O ₂	O ₃	O ₄
2	X ₁	X ₂	X ₃	X ₄	O ₁	O ₂	O ₃	O ₄
3	X ₁	X ₂	X ₃	X ₄	O ₁	O ₂	O ₃	O ₄
4	X ₁	X ₂	X ₃	X ₄	O ₁	O ₂	O ₃	O ₄

Para un adecuado análisis de resultados, se parte de la separación de la variación total en: a) la variación debida a las diferencias individuales, y b) la variación intra-sujeto. A su vez, puesto que todos los tratamientos son aplicados a todos los sujetos, la variación intra-sujeto, se divide también en: a) la variación debida a los tratamientos, y b) la variación debida al error. Por tanto, podríamos esquematizar este análisis de la variancia de la siguiente forma (Winer, 1971):

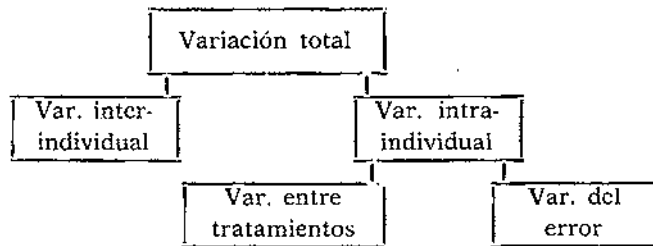


Fig. 11

Con este diseño disponemos de un procedimiento para llegar a un máximo control de una de las principales fuentes de variación: las diferencias inter-individuales. No obstante, no siempre es aconsejable su aplicación debido a una serie de dificultades que serán analizadas en el apartado siguiente.

5.2.4.b. Ventajas e inconvenientes

El diseño de medidas repetidas es sumamente útil, siempre y cuando, los efectos de los tratamientos tengan un efecto pasajero, y consecuentemente, puedan ser aplicados en un plazo de tiempo relativamente corto. Por lo general, su utilización es aconsejable, en psicología, en situaciones de tratamientos secuenciales, como en tareas de aprendizaje, así como, para el estudio del efecto de diversas drogas, siempre que su acción pueda ser eliminada en un tiempo breve.

Entre los inconvenientes más graves debemos señalar el ya conocido «error progresivo» o «efecto pro-activo de la práctica». Este efecto puede repercutir positiva o negativamente sobre los resultados. Es decir, el sujeto o bien puede acusar el cansancio motivado por la práctica repetitiva, o bien, la acción facilitadora del aprendizaje.

Ahora bien, en el caso que los diferentes tratamientos (como por ejemplo, el estudio del aprendizaje en función de la cantidad de ensayos o refuerzos) guarden entre sí una determinada proporción y consecuentemente, nos interese conocer el tipo de relación funcional existente entre las variables experimentales (función lineal, cuadrática, cúbica, etc.), podemos aplicar los resultados de este diseño el «análisis de tendencias», y probar hasta qué punto los tratamientos guardan entre sí una relación lineal.

A modo de punto final, diremos que es muy importante insistir de nuevo en el efecto de los tratamientos. Su acción debe ser rápida, sobre todo si se trata de drogas, para evitar el posible confundido entre ellas.

5.3. Experimentos factoriales

Los experimentos factoriales son aquellas disposiciones en las que se estudia, simultáneamente, la acción conjunta de dos o más variables independientes, llamadas, también, factores. Cada uno de estos factores suele tener varios valores, siendo el diseño factorial el conjunto de todas las posibles combinaciones entre los valores de ambas variables. De ahí que todo diseño factorial (en el caso que sea completo) se den tantos grupos como posibles combinaciones se puedan realizar entre los diversos tratamientos experimentales.

Son, por otra parte, notorias las ventajas que importa el uso de diseño factorial y que, brevemente, vamos a comentar. En primer lugar nos permiten un considerable ahorro de tiempo, coste y sujetos. Con un diseño factorial podemos obtener, con la misma cantidad de sujetos, doble información que la que se obtendría en caso de utilizar un diseño de factor único.

Vamos a ejemplificar este extremo mediante un ejemplo. Supongamos que tenemos dos variables independientes A y B, y queremos conocer hasta qué punto cada una de ellas afecta el fenómeno experimental. Si para ello utilizamos dos diseños clásicos de «dos grupos», uno para la variable A (con dos niveles: A_1 y A_2) y otro para la variable B (también a dos niveles: B_1 y B_2), y deseamos formar grupos de 20 sujetos, necesitaríamos un total de 80 sujetos, para llevar a cabo dicha investigación.

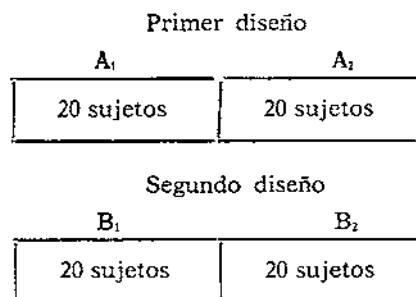


FIG. 12

Es decir, para probar por separado el efecto de cada una de las variables independientes (A y B), necesitaríamos un total de 80 sujetos, seleccionados al azar de una misma población. En cambio, si para este mismo experimento, utilizáramos un diseño factorial (2×2), la cantidad de sujetos quedaría reducida a la mitad.

Diseño factorial 2×2

A				
		A ₁ A ₂		
B	B ₁	10	10	n=20
	B ₂	10	10	n=20
		n=20	20	

FIG. 13

Con la misma cantidad de sujetos que utilizamos en diseños de un solo factor, para cada comparación (40 sujetos), podemos, mediante un diseño factorial, realizar las dos comparaciones. El ahorro de sujetos es notoria.

En segundo lugar, hemos de afirmar que con un diseño factorial no sólo ahorramos sujetos para la obtención de los mismos resultados, sino que además podemos calcular la posible interacción entre la variable A y B. Así, se puede probar que el efecto por separado de las variables sea nula, y que sin embargo, su interacción sea efectiva. En muchos casos, el experimentador puede estar mucho más interesado por el efecto de interacción, que por la posible actuación de cada una de las variables. Ello nos permite una mayor posibilidad de generalización, así como obtener una información mucho más valiosa, que lo que se conseguiría con diseños simples. En efecto, en el diseño factorial, en vez de tener neutralizadas una serie de variables sistemáticas extrañas, las introducimos en la situación experimental, con lo que el experimento es mucho más informativo. A su vez, podemos probar si el valor de una de las variables varía a medida que cambian las otras.

Por último, hemos de mencionar una tercera ventaja a favor de los diseños factoriales, sobre todo cuando es aplicado al estudio de los fenómenos conductuales. La conducta es una actividad compleja y se halla, por tanto, constantemente sometida a la acción de una gran variedad de factores. Mediante la utilización de diseños factoriales contamos con un procedimiento que nos va a permitir estudiar el cambio que experimenta la conducta en función de la acción conjunta de los factores. Esto nos permitirá estudiar tales fenóme-

nos desde una perspectiva más real que lo que se conseguiría con la utilización de diseños simples.

En suma, afirmaremos que el diseño factorial añade, junto a un positivo ahorro de sujetos y esfuerzo, un incremento del poder informativo de los resultados. Todo lo cual sitúa este tipo de diseños en un lugar privilegiado para la investigación de los fenómenos conductuales.

5.3.1. *Efectos factoriales: principales y secundarios*

Hemos afirmado antes que en todo diseño factorial se combinan dos o más variables o factores de dos o más valores. Cada una de las posibles combinaciones recibe el nombre de «tratamiento», y se aplica a una unidad o grupo experimental. De esta forma, cada grupo experimental representa una determinada combinación de dos o más valores.

El efecto de una variable independiente para cada uno de los valores de la otra, recibe el nombre de «efecto simple». Por el contrario, el efecto total de una variable independiente promediado a través de todos los valores de la otra, recibe el nombre de «efecto principal» de la variable. De ahí que, los efectos independientes de cada uno de los factores, reciban el nombre de «efectos principales».

En el caso que se estudie el efecto simple de una variable para cada uno de los valores de la otra, tendremos el «efecto de interacción» o «efecto factorial secundario». Es decir, si la acción de un factor sobre la variable dependiente varía de acuerdo con los valores que toma la otra. Si tenemos, por ejemplo, la variable independiente A, cuyo efecto varía según los valores que toma una segunda variable B, nos encontramos ante el «efecto de interacción de $A \times B$ ». Este mismo efecto puede darse para el caso inverso.

Los efectos de interacción en los que sólo intervienen dos variables reciben el nombre de «efectos secundarios de primer orden»; si intervienen tres variables, «de segundo orden», etc. (de primer orden, $A \times B$; de segundo, $A \times B \times C$, etc.).

5.3.2. *Diseño factorial «2×2»*

A fin de estudiar cómo se calculan los efectos principales y secundarios de un diseño factorial, vamos a presentar a continuación un diseño factorial simple: el diseño «2×2». En el diseño factorial «2×2» se combinan dos variables independientes, que podemos llamar A y B, siendo cada una de ellas a dos niveles (ésta es la razón por la que a estos diseños se los denomine «2×2» o «2²»; es decir, diseño de dos variables a dos niveles).

Vamos a utilizar las correspondientes letras minúsculas (*a* y *b*), para significar los diferentes valores de cada factor. Emplearemos, también, la nota-

ción de Yates (1937), para tener una mayor simplicidad en la presentación del diseño. Según Yates, la presencia de la correspondiente letra minúscula, representa el valor más alto de una variable; en cambio, su ausencia, el valor más bajo. Si tenemos, por ejemplo, la combinación «a», ésta denotará un tratamiento formado por el nivel más bajo de «B» (puesto que en el símbolo no aparece «b»), y el nivel más alto de «A». En el caso que se combinen los dos niveles bajos de «A» y «B», el tratamiento viene representado por el símbolo «(1)».

En la figura 14 presentamos una tabla correspondiente a un experimento factorial «2×2», en donde se hallan especificadas las diferentes combinaciones posibles entre los diversos niveles de «A» y «B».

Plan factorial 2×2
según la notación de Yates

		A	
		1	2
B	1	(1)	a
	2	b	ab

FIG. 14

Cada uno de los cuadrados o celdas de la tabla, representa una de las cuatro posibles combinaciones de los valores de las variables independientes, a los que asignaremos una cantidad «n» de sujetos.

Si simbolizamos el valor medio de las observaciones mediante la letra X , podemos calcular los efectos principales de «A» y «B», con la aplicación de las siguientes ecuaciones:

$$(1) \text{ Efecto principal de «A»} = 1/2 [(\bar{X}_{ab} - \bar{X}_b) + (\bar{X}_a - \bar{X}_{(1)})]$$

$$(2) \text{ Efecto principal de «B»} = 1/2 [(\bar{X}_b - \bar{X}_{(1)}) + (\bar{X}_{ab} - \bar{X}_a)]$$

También podríamos calcular estos mismos efectos, comparando las medias combinadas de los cuadrados (1) y b , con la combinación de medias en a y ab , para el efecto principal de «A»; siguiendo el mismo procedimiento para calcular el efecto principal de «B» (es decir, comparando la combinación de las medias (1) y a , con la combinación de las medias de b y ab).

$$(3) \text{ Efecto de interacción «A×B»} = 1/2 [(\bar{X}_a - \bar{X}_{(1)}) - (\bar{X}_{ab} - \bar{X}_b)]$$

Para calcular, pues, la interacción entre «A» y «B» se buscará en primer lugar el efecto simple de «A» para un valor dado de «B» (mediante la diferencia entre la media en «a» y «(1)», para el primer nivel de «B»), y a continuación, calcularemos el efecto simple de «A» para el segundo nivel de «B» (a través de la diferencia entre las medias en «ab» y «b»), y se constatarán ambos efectos. La diferencia entre este par de diferencias nos va a dar el valor de la interacción «A×B».

Se podría calcular también dicha interacción computando los efectos simples de «B» para los dos niveles de «A», según la siguiente ecuación:

$$(4) \text{ Efecto de interacción «B} \times \text{A»} = 1/2 [(\bar{X}_b - \bar{X}_{(1)}) - (\bar{X}_{ab} - \bar{X}_a)]$$

Tenemos, pues, dos expresiones —(3) y (4)— que nos permiten el cálculo de la interacción entre ambos factores, y se reducen en obtener la media de la diferencia de los efectos simples de un factor. Podríamos simplificar algebraicamente cada una de las expresiones anteriores de la forma siguiente:

$$(3a) \text{ Efecto de la interacción «A} \times \text{B»} = 1/2 [(\bar{X}_{(1)} + \bar{X}_{ab}) - (\bar{X}_a + \bar{X}_b)]$$

$$(4a) \text{ Efecto de la interacción «B} \times \text{A»} = 1/2 [(\bar{X}_a - \bar{X}_b) - (\bar{X}_{(1)} - \bar{X}_{ab})]$$

Por tanto, podemos concluir que la interacción «A×B» consiste en la comparación cruzada de dos pares de tratamientos con los otros dos; es decir, la comparación de los tratamientos que se hallan en las diagonales del plan experimental.

5.3.3. *Esquema básico para el tratamiento estadístico de los resultados de diseños factoriales*

En el apartado anterior hemos presentado una serie de comparaciones entre las medias de los diferentes tratamientos del plan experimental, lo que nos ha servido para la presentación y comprensión de los diferentes efectos factoriales. En la práctica, para el cálculo de tales efectos no suele procederse de la forma antes indicada, aunque ésta nos haya servido para analizar el procedimiento lógico para la estimación de tales efectos. En el caso de que nos interese conocer, de una forma más fiable y significativa, el efecto de las posibles variables sobre el fenómeno estudiado, así como su posible interacción, se suele utilizar el «análisis de la variancia», el cual constituye un procedimiento mucho más preciso para el estudio de todos los efectos factoriales.

El principio básico sobre el que se asienta el «análisis de la variancia» consiste en dividir la variación total de los resultados en dos componentes de variación: 1) el debido a los tratamientos experimentales (o variación entre

tratamientos), y 2) la variación debida al error (o variación intratratamientos).

A su vez, la «variación entre tratamientos» se puede dividir en una serie de componentes, según se agrupen los resultados en un sentido u otro. Estas diferentes agrupaciones de los datos, base de los componentes de variación entre grupos, se realizan de acuerdo con los diversos factores y de su interacción. Así concretamente, para un diseño factorial « 2×2 », con cuatro grupos experimentales, existen « $4 - 1$ » componentes «entre», a saber: a) el componente de variación debido al factor «A»; b) el componente de variación debido al factor «B», y c) el componente de variación debido a la interacción «A \times B». Y de la misma forma que la variancia «entre grupos» se divide en tres componentes (para un diseño factorial « 2×2 »), también los grados de libertad asociados a la misma se dividen en tres. Es decir, uno para cada componente.

Vamos a presentar a continuación una tabla que nos va a servir de esquema para el cálculo y estimación de las variaciones de cada uno de los componentes de variabilidad en lo que se divide la variancia total de los resultados.

Fuentes de variación	Suma de cuadrados	Grados G. libertad	Cuadrado medio	Razón variancia
Entre tratamientos	SC_x	« $k - 1$ »	$SC_x/k - 1$	
Factor A	SC_a	1	$SC_a/1 = a$	a/d
Factor B	SC_b	1	$SC_b/1 = b$	b/d
Interac. A \times B	SC_{ab}	1	$SC_{ab}/1 = c$	c/d
<u>Intra grupos (error)</u>	SC_e	« $k(n-1)$ »	$SC_e/k(n-1) = d$	
<u>Variancia total</u>	$SC_x + SC_e$	« $kn - 1$ »		

Una vez calculadas las sumas de cuadrados (SC), para los diversos componentes en que hemos dividido la variancia total de los datos, éstos se transforman en índices comparables entre sí, al ser divididos por sus respectivos grados de libertad. De esa forma los «SC» se transforman en «cuadrados medios» (CM). Cada uno de los CM, correspondientes a los componentes de variación, se comparan con el CM del error, para obtener de esta forma «el índice de variación» o «razón de la variancia», el cual nos va a permitir conocer el grado de significación de los resultados desde un punto de vista estadístico. Mediante estas razones de variación contamos con una base razonable para inferir hasta qué punto la acción de cada uno de los factores han ejercido una influencia sobre los resultados mayor de la que cabría esperar del azar.

5.3.4. Modelos de diseños factoriales

Al planificar un diseño experimental factorial solemos elegir una serie de valores para cada una de las variables independientes, de una forma aleatoria,

o al menos arbitraria. En efecto, si estamos interesados en conocer la influencia del «impulso-ansiedad» sobre un tipo de «rendimiento motor», usualmente se parte de unos valores dados como, por ejemplo, «alta y baja ansiedad», sin que sean elegidos de entre la amplia gama de niveles impulsionales. Es ésta, al menos, la forma con la que usualmente se procede.

Por tanto, cuando establecemos, según algún criterio apriorístico, los niveles de nuestras variables independientes, el diseño factorial recibe el nombre de «*modelo fijo*». En el caso que los valores de las variables independientes se hayan elegido al azar, el diseño factorial toma el nombre de «*modelo al azar*». Puede darse también el caso intermedio. Es decir, que una de sus variables tengan los valores elegidos al azar, y, en cambio, la otra según un criterio arbitrario o fijo. En este último caso tendríamos un diseño factorial de «*modelo mixto*».

A continuación vamos a estudiar estos modelos para un diseño factorial de dos factores. Nos fijaremos, fundamentalmente, en los valores esperados para uno de sus componentes de variación. Del análisis de los diversos componentes de variación, podremos llegar a establecer el «adecuado término de error» para la correspondiente prueba de significación.

5.4.3.a. *Modelo factorial fijo*

Partamos del supuesto que un investigador haya elegido, de un modo arbitrario, los valores de sus dos variables independientes. Si estuviese interesado, por ejemplo, en estudiar el efecto de la «ansiedad» y el «tipo de práctica» sobre el «aprendizaje de pares asociados», podría elegir para la ansiedad los valores «alto» y «bajo», y para el tipo de práctica, las modalidades «masiva» y «distribuida». Puesto que parte de unos valores convencionales, podemos afirmar que está utilizando un «modelo factorial fijo».

Vamos a presentar a continuación la expresión matemática en la que se recogen los diversos componentes de variación de dicho modelo.

$$X = m + a + b + ab + e$$

Siendo «X» una observación o dato cualquiera de nuestro experimento; «m» simboliza la media general de todas las observaciones; «a» viene a representar el componente de variación debido a los valores del tratamiento «A»; «b», el componente de variación debido a la variable «B»; «a» es el componente de variación debido a la interacción de los factores independientes, y «e» es el componente que sumaría la acción de todas las fuentes de error.

En el caso que la «hipótesis nula» fuera verdadera, la suma de los efectos debidos a «a», «b» y «ab» quedarían anulados, quedando como única fuente de variación «el error experimental». En caso contrario, debe añadirse al com-

ponente de variación debido al error la variancia producida por un determinado factor específico.⁶

En la tabla siguiente presentamos los valores esperados en la estimación de la variancia, en el supuesto de que la H_0 e H_1 sean respectivamente verdaderas.

Fuente de variación	Variancia esp. bajo H_0	Varianza esp. bajo H_1
Tratamiento A	σ^2	$\sigma^2 + nq\sigma^2$
Tratamiento B	σ^2	$\sigma^2 + np\sigma^2$
Interacción A \times B	σ^2	$\sigma^2 + n\sigma^2$
Intragrupo (error)	σ^2	σ^2

(siendo «n» el número de sujetos por grupo; «p», la cantidad de valores del tratamiento «A», y «q», la cantidad de valores del tratamiento «B»).

En este modelo, puesto que cada elemento está formado por dos componentes de variación, bajo el supuesto de la hipótesis alternativa como verdadera tendremos: a) por una parte el componente debido a la variancia del error, y b) el componente de variación debido a un determinado factor, el adecuado término de error será la «variancia intragrupo». Si el efecto de cada uno de los componentes específicos de variación fueran nulos, sólo compararíamos las «variancias de error». En este caso, la «razón de variancia» sería igual a 1 para cada una de las comparaciones, y consecuentemente concluiríamos con la aceptación de la hipótesis nula.

5.4.3.b. Modelo factorial al azar

Cuando el experimentador ha elegido los valores para sus variables independientes de una forma totalmente aleatoria, tenemos un «modelo factorial al azar». Un ejemplo para este modelo podría ser un experimento en el que se estudiaran el efecto de la «técnica de diagnóstico» y el del «procedimiento terapéutico» sobre el «mejoramiento de los enfermos mentales». Si el investigador no muestra preferencia alguna por un tipo determinado de diagnóstico o de terapia, puede elegir al azar entre una larga lista de métodos de diagnóstico y procedimientos terapéuticos, dos de ellos respectivamente.

6. Se parte del supuesto, para un tipo de modelo matemático lineal, que la distribución del error es normal y su efecto es independiente de los componentes restantes.

Hemos de indicar que para un modelo al azar, los componentes de variación que afectarían a cada una de las observaciones siguen también el modelo matemático lineal, como el presentado antes.

Bajo el supuesto de que la «hipótesis nula» sea verdadera, la esperanza de los efectos de variación debidos a los dos tratamientos y a su interacción sería igual a cero. Por el contrario, bajo el supuesto de la «hipótesis alternativa» como verdadera, puesto que los valores de las variables independientes han sido elegidos al azar, en la estimación de la variación para los respectivos tratamientos, tendríamos tres componentes: a) el componente debido al error; b) el componente debido a la interacción, y c) el componente específico debido a cada uno de los factores independientes. Ésta es una de las razones por las que este modelo sea conocido también con el nombre de «modelo del componente de la variancia», ya que éste aparece en cada uno de los efectos factoriales.

En la tabla que vamos a presentar a continuación damos los valores esperados de las variancias para cada uno de los efectos factoriales, bajo la suposición de la hipótesis nula y la alternativa como verdadera.

Fuente de variación	Variancia esp. bajo H_0	Varianza esp. bajo H_1
Tratamiento A	σ_e^2	$\sigma_e^2 + n\sigma_{ab}^2 + nq\sigma_u^2$
Tratamiento B	σ_e^2	$\sigma_e^2 + n\sigma_{ab}^2 + np\sigma_b^2$
Interacción A×B	σ_e^2	$\sigma_e^2 + n\sigma_{ab}^2$
Intragrupo (error)	σ_e^2	σ_e^2

En el supuesto de la hipótesis nula como verdadera, se espera que la variancia debida a la interacción, así como la variancia debida a los factores independientes sean iguales a cero. Por esa razón, al probar la significación del efecto de las variables independientes, puesto que se trata de un modelo al azar, y, consiguientemente, tales efectos estarán afectados por la variancia de la interacción, se elegirá como adecuado término de error (elemento de contrastación para probar la acción de los tratamientos) la «variancia intragrupo más la variancia de interacción». Es decir, la suma de cuadrados correspondiente a la interacción. Por otro lado, para probar el efecto de interacción, se tomará como término de error o contrastación «la variancia intragrupo o del error».

Podemos, por tanto, constatar que para este modelo (llamado también componente de la variancia) los efectos de los tratamientos se hallan afectados por la variancia de la interacción. Este componente de contaminación de

la variancia debida a los tratamientos tiene su origen en el hecho de que los valores de las variables independientes hayan sido elegidos al azar. Y es, precisamente, dicha selección al azar la causa de la interacción entre los factores, ya que para su anulación deberían tenerse en cuenta todos los valores. Si en el ejemplo propuesto se hubiese elegido al azar uno de los posibles diagnósticos como, por ejemplo, la neurosis, es posible que ésta puede ser más o menos curable por alguno de los tratamientos elegidos al azar. Esto hace que dicha variable se contamine y se produzca, consecuentemente, un efecto de interacción. Y ésta es la razón por la que dicho efecto deberá tenerse en cuenta en el momento de analizar las variancias debidas a los tratamientos.

5.4.3.c. Modelo factorial mixto

Un caso mucho más corriente suele presentarse en investigación experimental, sobre todo en aquellas situaciones en que uno de los factores independientes tiene sus niveles elegidos al azar, y, en cambio, el segundo tiene sus valores determinados arbitrariamente debido a exigencias teóricas. Estamos, pues, ante un modelo factorial mixto. Este modelo suele emplearse, regularmente, para aquellos diseños en que una de las variables independientes es el propio sujeto experimental. En este caso se procedería de la siguiente forma. Se elegirían al azar una muestra de sujeto de entre una población dada. A continuación someteríamos esta muestra a una serie de condiciones experimentales previamente seleccionadas. De esta forma, tendríamos un diseño factorial mixto con una variable elegida al azar, y una segunda mediante un criterio arbitrario o fijo. La tabla siguiente representa la esperanza de valores en la estimación de las variancias bajo la suposición de la hipótesis nula y bajo la suposición de la hipótesis alternativa.

Fuentes de variación	Variancia esp. bajo H_0	Variancia esp. bajo H_1
Tratamiento A (v. fija)	σ_c^2	$\sigma_c^2 + n\sigma_{cb}^2 + nq\sigma_a^2$
Tratamiento B (v. al azar)	σ_c^2	$\sigma_c^2 + np\sigma_b^2$
Interacción A×B	σ_c^2	$\sigma_c^2 + n\sigma_{ab}^2$
Intragrupo (error)	σ_c^2	σ_c^2

Fijándonos en la tabla anterior, comprobaremos que la variancia esperada para la variable fija «A» se halla afectada por un componente de variación de interacción. Dicho componente es producido por la selección al azar de los

valores de «B». De ahí que el adecuado término de error para probar la significación de los efectos de «A» es la «variancia del error más la variancia de la interacción». Bajo la hipótesis nula, se supone que el efecto de esta variación es igual a cero. Pero cuando tratamos de probar «A», bajo la suposición de la hipótesis alternativa, se deberá neutralizar el componente de interacción al dividir su valor esperado por el de la interacción.

Para probar «B», puesto que no se halla afectada por la variancia de la interacción, el término adecuado de error será únicamente «la variancia del error». Término que nos va a servir, a su vez, para probar la significación de la interacción.

Por último, hemos de tener en cuenta que el valor esperado de «B» no se halla afectado por el componente de interacción, ya que, como hemos indicado, los valores de «A» no han sido elegidos al azar.

5.4. Principio de la confusión

Suele ocurrir con frecuencia, sobre todo en psicología, que al definir una variable mediante la selección de grupos distintos de sujetos, como por ejemplo el «nivel socioeconómico», y se intenta estudiar su influencia sobre una variable de medida, como por ejemplo «preferencias vocacionales», ésta acusará la influencia de otros factores como «las posibilidades ocupacionales futuras», «el escaso nivel cultural recibido», etc., afectándola sistemáticamente. Como es obvio, puede esperarse que sujetos pertenecientes a estratos sociales inferiores consideren sus posibilidades de trabajo, para determinadas ocupaciones, más limitadas. Y, por tanto, ello puede influir decisivamente en su preferencia vocacional, a veces más incluso que el propio nivel socioeconómico.

Lo que acabamos de afirmar viene a ser una ilustración de una variable independiente (nivel socioeconómico) totalmente contaminada. Los metodólogos suelen denominar tales variables «variables confundidas». En efecto, siempre que el cambio de valores de un factor vaya acompañado sistemáticamente por el cambio de otro factor no controlado, tendremos un «efecto de confusión». De hecho, uno de los objetivos básicos del control consiste, precisamente, en evitar este tipo de efectos.

Ahora bien, muchas veces nos encontramos en situaciones en las que es necesario aplicar diseños factoriales con un gran número de tratamientos (de 12 a 16 combinaciones). Ello exigiría la utilización de una gran cantidad de sujetos, si realmente nos interesase hacer una réplica completa del experimento. Por esta razón, muchas veces el experimentador opta por una reducción del tamaño de la réplica. Para ello se suele sacrificar, deliberadamente, alguna interacción a fin de conseguir una mayor precisión en la estimación de los efectos principales.

Vamos a ilustrar este principio a través de un diseño muy sencillo. Si bien el principio de confundido no suele aplicarse a diseños sencillos, aquí lo utilizaremos a efectos metodológicos.

Tomemos, por ejemplo, un diseño factorial «2×2» y dividámoslo de la siguiente forma (para ello emplearemos la notación de Yates):

Bloque I	Bloque II
(1)	a
ab	b

Para el cálculo del efecto principal de «A», buscaríamos la diferencia entre las medias en «ab» y «(1)» para el primer bloque, y la diferencia entre las medias en «a» y «b» para el segundo, para sumarlas a continuación. De esta forma, los efectos de «A» estarían equilibrados en relación a los efectos de «B».

$$(5) \text{ Efecto principal de «A»} = 1/2 [(\bar{X}_{ab} - \bar{X}_{(1)}) + (\bar{X}_a - \bar{X}_b)]$$

y para el cálculo del efecto principal de «B» aplicaríamos:

$$(6) \text{ Efecto principal de «B»} = 1/2 [(\bar{X}_{ab} - \bar{X}_{(1)}) + (\bar{X}_b - \bar{X}_a)]$$

para el cálculo del efecto de interacción «A×B» tendríamos:

$$(7) \text{ Efecto de interacción «A×B»} = 1/2 [(\bar{X}_{ab} - \bar{X}_b) - (\bar{X}_a - \bar{X}_{(1)})]$$

Esta última ecuación es matemáticamente equivalente a:

$$(8) \text{ Efecto de interacción «A×B»} = 1/2 [(\bar{X}_{ab} + \bar{X}_{(1)}) - (\bar{X}_a + \bar{X}_b)]$$

De la ecuación (8) podemos deducir que para el cálculo de la interacción «A×B», hemos buscado el total de las medias de observaciones del bloque I y le hemos restado el total de medias del bloque II, buscando a continuación su valor promedio. De esta forma podemos comprobar que la comparación de tratamientos que hacemos para el cálculo de la interacción coincide con la comparación de los dos bloques. Así, podemos afirmar que la interacción «A×B» queda confundida con el efecto del bloqueo.

Es interesante tener presente que el cálculo de los efectos factoriales para un tipo de diseño «2ⁿ» (de «n» factores a dos niveles respectivamente), consiste en la comparación de una mitad de las medias de las observaciones con la otra mitad. Es decir, siempre una mitad se presentará con el signo más, y la otra con el signo menos. Así, si tenemos un experimento factorial «ABCD» (con dos valores respectivamente: 2⁴), su ejecución completa supondría el empleo de dieciséis condiciones o combinaciones de tratamiento.

Supongamos que hemos asignado «n» sujetos al azar a cada uno de los dieciséis grupos. Por otra parte, la tarea a realizar es bastante compleja y deberá ser realizada a lo largo de varios días. Ésta implicará el que algunos

sujetos realicen la prueba por la mañana y otros por la tarde. En el caso que quisiéramos controlar la variable «mañana-tarde», deberíamos aleatorizar los sujetos de tal manera que cada uno de ellos realizara las pruebas, en los diferentes días, tanto por la mañana como por la tarde. Con ello aumentaríamos considerablemente el error experimental, con lo que, automáticamente, nuestro diseño perdería sensibilidad. Para evitar esto, podríamos formar dos bloques de sujetos más homogéneos, de forma que los que pertenezcan a un bloque realizaran las pruebas sólo por las mañanas, y los pertenecientes al otro sólo por las tardes. En este caso tendríamos en cada bloque la mitad de sujetos, y consecuentemente, tendríamos en cada bloque menos sujetos de los que serían necesarios para realizar una réplica completa del experimento.

Para solucionar este problema, podríamos proceder de la siguiente forma. Si partimos del supuesto de que la variable «mañana-tarde» no es muy importante para nuestro experimento, a fin de obtener conclusiones válidas, la podemos sacrificar confundiendo con la interacción de orden superior. En nuestro caso sería $A \times B \times C \times D$. Y ello, como hemos apuntado, debido a la escasa relevancia que tendría para nuestro trabajo.

Por lo tanto, partiendo del propósito básico de sacrificar la interacción $A \times B \times C \times D$, podemos dividir el experimento en dos bloques, de forma que cada bloque incluya la mitad de las condiciones (es decir, ocho combinaciones de tratamiento). Así, un bloque estaría formado por aquellos grupos que en la comparación para la interacción aparecen con el signo «más», y el segundo bloque con los grupos que toman el signo «menos». Existe una regla, de sencilla aplicación, que nos permite conocer el signo que va a tener cada grupo para cualquier tipo de interacción en un diseño de dos niveles por factor (Cochran y Cox, p. 188). Consideremos, por ejemplo, la interacción cuádruple de nuestro ejemplo:

$$(9) \quad A \times B \times C \times D = 1/2^{n-1} (a-1)(b-1)(c-1)(d-1).$$

Desarrollando este polinomio y agrupando los términos positivos a un lado y los negativos a otro, podemos formar dos bloques de la forma siguiente:

Cálculo de la interacción $A \times B \times C \times D$
Comparación

+	-
abcd	abc
ab	abd
ac	acd
ad	bcd
bc	a
bd	b
cd	c
(1)	d

La comparación de los grupos para el cálculo de la interacción cuádruple viene esquematizada en la tabla anterior. Si aprovechamos dicha disposición, a un lado los grupos con signo más y al otro los grupos con signo menos, para formar dos bloques de ocho tratamientos cada uno, y hacemos pasar las pruebas para los grupos del bloque con signo positivo por la mañana, y para los del signo negativo por la tarde, habremos confundido la variable «mañana-tarde» con la interacción $A \times B \times C \times D$. Esto significará que en nuestro análisis de resultados no podremos separar ambos efectos. Con ello hemos sacrificado una información al confundir la interacción con una variable. No obstante, ello nos ha permitido realizar un experimento con un considerable ahorro de sujetos.

Como punto final, señalemos que la práctica del confundido es aconsejable siempre que la identificación de una variable con una interacción no sea de gran interés para nuestros resultados experimentales. Es decir, siempre que el efecto de dicha interacción pueda ser despreciado.

6. Resumen

En este escrito hemos intentado presentar la problemática que entraña la utilización del diseño experimental para las ciencias de la conducta. Si bien, en principio, todo plan experimental encierra ya de por sí un determinado diseño, no por ello debemos de ignorar la existencia de modelos de diseños, cuya utilidad y eficacia se ha demostrado a lo largo de los últimos años.

Nuestra intención no ha sido la de desarrollar exhaustivamente la compleja temática que implica la utilización de un determinado diseño. Solamente hemos pretendido presentar un cuadro básico de los principales diseños de experimentación que, a nuestro juicio, pueden ser empleados con éxito en la investigación psicológica.

RÉSUMÉ

Nous avons essayé de présenter par ce texte la problématique que comprend l'usage du design expérimental pour les sciences de la conduite. Même si en principe tout plan expérimental comporte déjà de lui-même un design déterminé, ce n'est pas pour cette raison que nous devrions ignorer l'existence de certains modèles de design, dont l'utilité et l'efficace sont reconnues depuis ces dernières années.

Nous n'avons pas eu l'intention de développer de forme exhaustive la complexe thématique de l'utilisation d'un certain design; nous désirions seulement offrir une référence de base des principaux designs d'expérimentation, qui selon nous, peuvent être employés avec succès par la recherche psychologique.

SUMMARY

In this paper we have tried to present the implications, for the behavioural sciences, of the use of experimental design. Although, essentially, any experimental plan in itself presupposes a certain design, yet we must not ignore the existence of design patterns whose utility and efficiency have been proved over the last years.

It has not been our intention to develop exhaustively the complex problems implied in the utilization of any specific design. We only wish to present a basic system of the principal experimentation designs which, in our opinion, could be successfully employed in psychological research.

BIBLIOGRAFIA

- BRADY, J. V.: «Ulcers in "executive" monkeys». *Scient. Am.*, 199, 95-100 (1958).
- ANDREWS, T. G. (ed.): *Methods of Psychology*. J. Wiley, Nueva York, 1948 (existe trad. francesa en P.U.F.).
- BORING, E. C.: «The nature and history of experimental control». *Amer. J. Psychol.*, 67, 573-589 (1954).
- BRUNSWIK, E.: *Systematic and representative Design of psychological Experiments*. Univ. of California Press, Berkeley, 1947.
- BRUNSWIK, E.: *Perception and the representative Design of psychological Experiments*. Univ. of California Press, Berkeley, 1956.
- CAMPBELL, D. T.: «Factors relevant to validity of experiments in social settings», *Psychol. Bull.*, 54, 297-312 (1957).
- COCHRAN, W. C.; COX, G. M.: *Experimental Designs*. J. Wiley, Nueva York, 1957 (existe trad. cast. en Ed. Trillas).
- COX, D. R.: *Planning of experiments*. J. Wiley, Nueva York, 1958.
- EDWARDS, A. L.: *Experimental Design in psychological Research*. Rinehart, Nueva York, 1960.
- FISHER, R. A.: *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935.
- GORMEZANO, I.; MOORE, J. W.; DEAUX, E.: «Yoked comparisons of classical and avoidance eyelid conditioning under three UCS intensities», *J. of Exp. Psychol.*, 64, 551-552 (1962).
- HAMMOND, R.: «Representative vs. systematic design in clinical psychology». *Psychol. Bull.*, 51, 150-159 (1954).
- LINDQUIST, E. F.: *Design and analysis of experiments in psychology and education*. Houghton Mifflin Comp., Boston, 1953.
- MATALON, B.: «La logique des plans d'expérience», en Lemaine, G., y Lemaine, J. M. (eds.): *Psychologie sociale et experimentation*. Mouton y Bordes, Paris, 1969.
- PAYNE, S. L.: «The ideal model for controlled experiments». *Publ. Opin. Quart.*, 11, 557-562 (1951).
- SOLOMON, H. L.: «An extension of the control group design», *Psychol. Bull.*, 46, 137-150 (1949).
- SOLOMON, R. L.; LASSAC, M. S.: «A control group design for experimental studies of developmental processes». *Psychol. Bull.*, 70, 145-150 (1968).
- REUHLIN, M.: «Utilisation en psychologie de certains plans d'expérience», *Ann. Psychol.*, 53, 58-81 (1953).
- UNDERWOOD, B. J.: *Psicología experimental*, Trillas, Méjico, 1972 (2.ª ed. inglesa, 1966).
- WINER, B. J.: *Statistical principles in experimental design*, McGraw-Hill, Nueva York, 1971 (2.ª ed.).