

CONDICIONES CIENTIFICAS DE UN EXPERIMENTO

Luis Valero Aguayo

Depto. Personalidad, Evaluación y Tratamiento Psicológicos
Facultad de Psicología, Málaga

Introducción. Cualquier estudio experimental -sea de grupos o de caso único- requiere un conjunto de condiciones que aseguren en su cumplimiento la certeza de los resultados obtenidos y de las inferencias que se realizan sobre los datos. Estas condiciones giran en torno a los conceptos de validez interna y externa de un experimento.

Estos conceptos se fundamentan en el trabajo de Campbell y Stanley (1973), y reelaborado en Cook y Campbell (1979), según los cuales un experimento con validez interna sería aquél en que podría eliminarse otras explicaciones alternativas a los mismos hechos, de forma que se podrían afirmar las relaciones causales entre el tratamiento y las medidas tomadas como debidas exclusivamente a las variables manipuladas.

Mientras que la validez externa sería la extensión con que los hallazgos de un experimento podrían ser generalizados a diferentes sujetos, situaciones, experimentadores o terapeutas.

Lo que diferencia a un experimento de otros métodos de obtención de datos es la modificación sistemática de una VI por parte del clínico (el tratamiento). Sólo esto permite el control de otras variables de interferencia sobre la VD (la conducta problemática), y posibilita una interpretación de los resultados en el sentido de una auténtica dependencia (Campbell y Stanley, 1973).

Experimento y cuasi-experimento

Esta distinción procede del texto clásico de Cook y Campbell (1979) donde se hablaba de "verdadero" experimento como aquél en que el experimentador puede asignar aleatoriamente una VI a un grupo determinado, si no puede asignar los sujetos a una condición experimental o -lo que es lo mismo- asignar un tratamiento a un grupo, sería un cuasi-experimento. En este último caso la falta de asignación aleatoria no permitiría inferir relaciones causales entre el tratamiento y los cambios en la VD.

Sin embargo, la estrategia de aleatorización no se aplica exclusivamente a la distribución de grupos, también puede ocurrir en diseños de caso único. En éstos últimos, el experimentador puede controlar la ejecución o no de una VI determinada en un momento determinado. En este sentido, podría haber experimentos y cuasi-experimentos con cualquier tipo de diseño entregrupos o intrasujeto.

Lo que marcaría el carácter experimental de un determinado estudio sería el control que puede imponer el experimentador a sus variables, y la seguridad que puede tener sobre las inferencias o conclusiones extraídas de los datos. El grado con que el experimentador pueda eliminar diversos factores que puedan afectar a los resultados, indicaría la extensión con que pueda realizar afirmaciones sobre relaciones causales entre las variables manipuladas (Cook y Campbell, 1979; Kiesler, 1981). Así, la tarea inicial de toda investigación será intentar disminuir los factores que afectan a la validez interna de un experimento, que limitarían lógicamente la seguridad sobre las conclusiones (Barlow y Hersen, 1988; Johnston y Pennypacker, 1980; Kazdin, 1982).

Validez interna del experimento

La validez interna se refiere a la exactitud científica de un estudio, puede afirmarse cuando el resultado de ese estudio puede atribuirse exclusivamente a la VI que se manipuló, y los resultados obtenidos no podrían ser explicados empíricamente por ningún otro tipo de variable. Literalmente, es la "base mínima sin la cual un experimento no resulta interpretable" (Campbell y Stanley, 1973, pag. 5). Cualquier variable -diferente a la VI- que pueda explicar los resultados de un experimento constituye una amenaza a la validez interna de un experimento (Kazdin, 1981). Entre esas amenazas Campbell y Stanley (1973) han delimitado las siete siguientes:

1. Historia. La ocurrencia de acontecimientos externos que afecten al comportamiento del individuo, es decir, cualquier tipo de hecho que le ocurra en su vida diaria podría afectar a la VD. Si esos acontecimientos ocurren, y el clínico o experimentador no los detecta, podría atribuir la mejoría a su tratamiento, cuando en realidad podría ser efecto de una llamada telefónica de un amigo, un programa de TV que haya visto, el inicio de una nueva relación sentimental, un aumento de sueldo, un cambio de trabajo, o cualquier otro cambio en su vida habitual.

2. Maduración. Se refiere a los cambios ocurridos en el individuo debidos a la propia maduración biológica o cualquier otro proceso de interacción que cambien a lo largo del tiempo. Por ejemplo, un clínico puede atribuir una mejora en el rendimiento de un niño cuando ese efecto podría deberse también a la maduración del niño o la mayor complejidad de sus interacciones al aumentar en edad.

3. Prueba. Los datos obtenidos con una primera prueba pueden cambiar al repetir la aplicación de esa misma prueba, por efecto de la práctica o el recuerdo. Aunque no se produzca ninguna intervención, la repetición de las pruebas de evaluación puede mostrar cambios. Por ejemplo, la aplicación de un cuestionario sobre ansiedad puede mostrar cambios entre una y otra aplicación, aunque no se haya iniciado el tratamiento.

4. Instrumentación. También se incluyen en este tipo de amenazas los errores de instrumentación, o errores en los propios procedimientos de evaluación. Un aparato no calibrado, con errores no sistemáticos, o procedimientos con poca fiabilidad, no validados, etc., impide llegar a conclusiones seguras sobre los resultados. Por ejemplo, una balanza de pie para control de peso que no esté bien calibrada, o un registro

portátil de la tasa cardíaca al que le fallen las pilas, pueden introducir errores sistemáticos que confundan los resultados.

5. Regresión estadística. Es un concepto matemático y se refiere al efecto de las puntuaciones extremas que tienden a "regresar a la media" de grupo. Cuando en la VD parecen existir efectos de techo o suelo, es decir, las puntuaciones son extremas y poco podrían aumentar o disminuir, respectivamente, entonces la tendencia en la siguiente medición es a obtener datos más cercanos a media, o al menos más estables, y no reflejarían el posible cambio de la VI. Por ejemplo, en una situación clínica un cliente podría obtener una puntuación excesivamente baja en un cuestionario o test de depresión, es probable que en la siguiente medición las puntuaciones tiendan a ser algo más elevadas, y entonces el clínico atribuya ese cambio a su psicoterapia.

6. Selección. Se refiere a las diferencias sistemáticas en la comparación de grupos que pueden ser resultado de una selección sesgada, no completamente aleatoria, en la que persiste alguna diferencia específica entre los grupos que podría explicar también los resultados. Por ejemplo, ocurriría si en un estudio sobre conductas depresivas los dos grupos proceden de instituciones sanitarias o tipos de consultas clínicas diferentes.

7. Mortalidad experimental. Es la pérdida de sujetos a lo largo de una experimentación con grupos, que hace cambiar la composición o características de los grupos comparados. Por ejemplo, cuando en un estudio sobre control de sobrepeso en uno de los grupos los clientes van abandonando progresivamente el tratamiento o dejan de acudir a consulta.

Posteriormente, se han añadido otras amenazas más específicas:

8. El efecto de la información puede hacer cambiar el comportamiento del individuo bajo experimentación (**efecto Harthorwe**), saber que está siendo observado o conocer que es objeto de experimentación puede ser suficiente para provocar cambios. En la psicoterapia este efecto es aún mayor por las propias expectativas que el sujeto suele llevar a la situación, y esas esperanzas de cambio pueden afectar a la validez interna (Wilkins, 1977). Como controles se utilizan el "efecto placebo" (O'Leary y Borkovec, 1969), es decir, se administra un sistema de tratamiento similar al experimental, pero que en absoluto tiene efecto comprobado. Este efecto aparece cuando hay factores no conocidos o situacionales en el contexto de la experimentación, y de hecho se utilizan situaciones "simuladas", contextos de psicoterapia sin la VI relevante, o instrumental similar al experimental.

9. Otro efecto de la información es cuando el terapeuta o experimentador conoce los posibles efectos del tratamiento, o espera de antemano unos resultados específicos (**efecto Rosenthal**, 1966). En estos casos, un adecuado control lo constituye el denominado "doble ciego", según el cual ni el sujeto experimental ni la persona que administra el tratamiento conocen qué tratamiento se administra (control o experimental) ni los supuestos efectos que debería provocar. Es especialmente relevante que la persona que realiza las evaluaciones, los registros, o la recogida de información, desconozca los tipos de tratamiento administrados a cada sujeto; así se aseguraría la independencia entre las medidas y la aplicación de la VI.

Cook y Campbell (1979) han enumerado también una serie de sesgos que pueden afectar a las conclusiones o inferencias obtenidos de un experimentos. Algunos de

estos factores pueden actuar en conjunto, y según el objeto de estudio, a veces pueden ser opuestos y el experimentador no tiene más alternativa que controlar tantos como pueda para poder llevar a conclusiones válidas. En casi todos ellos coinciden con los anteriores, pero añaden otras amenazas:

10. Interacciones de múltiples factores con la selección. Sería la ocurrencia simultánea de varios factores que afectasen a los resultados, por ejemplo, la historia y la maduración con sujetos con historias de desarrollo diferentes, o de maduración y selección que produciría grupos diferenciales a priori, o la instrumentación y la selección que ocurriría cuando las puntuaciones de las pruebas son diferentes para grupos con características diferentes.

11. Ambigüedad sobre la dirección de la influencia causal. Ocurriría cuando no es posible discernir -sin información sobre relaciones temporales- qué variable causa el efecto en la otra. Cuando se miden múltiples comportamientos, a veces no queda claro si A causa B o B causa A.

12. Difusión del tratamiento. Es posible que ocurra intercambio de información sobre los procedimientos, condiciones u objetivos de una investigación a través de sujetos en varios grupos. La validez dejaría bastante que desear pues no habría independencia entre los grupos a comparar.

13. Tratamientos de igualación o compensación. Ocurriría cuando se realizan comparaciones respecto a grupos con tratamientos compensatorios, a los que se aplica servicios o terapias con algún efecto, aunque sea mínimo, pues se considera inaceptable éticamente dejar a un grupo sin tratamiento alguno. El problema ocurriría porque, en estos casos, sería difícil llegar a afirmar la causalidad y eficacia de un único programa de tratamiento.

14. Reactividad en el grupo control. Hace referencia a efectos de rivalidad y/o desmoralización en los sujetos del grupo control. Cuando los sujetos conocen su pertenencia a un determinado grupo, que recibirá tratamientos con resultados menos deseables, esta información puede actuar para cambiar su comportamiento. Por lo que los sujetos podrían actuar por "rivalidad" o "dejadez" si conocen los resultados esperados de ese experimento. Estos efectos de reactividad impedirían conocer el efecto específico de un determinado tratamiento, independiente de esa reactividad.

A éstos habría que añadir una última amenaza que ocurriría cuando se trata de probar la eficacia de dos o más tratamientos:

15. Interferencia de tratamientos múltiples. Constituye un sesgo desde el momento en que se trate de probar dos o más tratamientos, y es fundamental su control en los diseños de caso único de tratamientos alternos. Habría tres efectos fundamentales a controlar para que no se dé esa interferencia (Barlow y Hayes, 1979; Ullman y Sulzer-Azaroff, 1975):

- **Efectos de orden.** Se refiere al posible cambio de resultados si el orden o secuencia de los tratamientos fuese diferente, controlado mediante secuencias aleatorias o semi-aleatorias de las condiciones experimentales.

- **Efectos residuales o acumulativos.** Hace referencia a los efectos de inducción que un tratamiento podría tener sobre otro, bien en sentido positivo o negativo respecto al

efecto real del tratamiento que se aplicase en segundo lugar. La investigación ha demostrado que pueden controlarse esos efectos mediante la alternancia rápida de las condiciones y control estricto de la situación estimular de experimentación.

- **Efectos de alternancia.** Se refiere a la capacidad del diseño para poder diferenciar los efectos de dos o más tratamientos, y no se produzca confusión con los efectos aislados de cada uno de ellos. Aquí es importante controlar las condiciones de aplicación, alternancia, secuencia y duración de las situaciones experimentales.

Validez externa de un experimento

La validez externa se refiere al grado en que los descubrimientos de una investigación pueden generalizarse a situaciones, personas o lugares distintos a los utilizados en ese estudio (Campbell y Stanley, 1973). En la investigación clínica la validez externa también es fundamental, por cuanto interesa generalizar los resultados al mayor número de clientes, problemas, situaciones o terapeutas posibles.

Campbell y Stanley (1973) definieron cuatro tipos de amenazas a la validez externa:

1. Reactividad a la prueba. Es decir, los cambios que puede mostrar el individuo por la aplicación de una prueba de evaluación, pues su comportamiento podría ser diferente si no se aplicase esa prueba. Por ello, sería difícil generalizar, por ejemplo, de un estudio sobre habilidades sociales en el que se utilice una situación simulada a otra en que no se realice ese tipo de pruebas.

2. Interacción de la selección con la VI. Supone que haya algunos efectos de la selección de los grupos, de forma que impida la generalización a otros grupos en que no haya ocurrido esa selección inicial, los sujetos de uno de los grupos podría mostrar efectos de reacción diferenciales a la manipulación de la VI. Por ejemplo, sería difícil generalizar a partir de un estudio sobre tabaquismo con un grupo compuesto por sujetos voluntarios altamente motivados para dejar de fumar, a otras personas que no muestren esa voluntariedad.

3. Reactividad a la situación experimental. Los sujetos suelen cambiar su comportamiento cuando se encuentran ante una situación estructurada, que exige un rendimiento determinado, con una gran directividad; frente a situaciones más naturales, donde no hay altas exigencias de la situación. Se han definido distintos tipos de reactividad a la situación experimental, por un lado, la reactividad observacional, que se produce cuando el sujeto cambia su comportamiento por el hecho de ser actor de una investigación y tiende a reaccionar más a las características generales de la investigación que a las VI efectivamente manipuladas. Por otro, el efecto de predisposición o sensibilización del sujeto a tratamientos posteriores, es decir, si el sujeto no es novato en las tareas experimentales puede que reaccione de forma diferente a otro que sí lo sea, y ello dificulte la generalización. Y, además, el denominado efecto de novedad, la reacción del sujeto a cambiar inmediatamente tras la introducción del tratamiento, sea cual sea éste, lo que impediría la generalización de esos resultados a largo plazo.

4. Interferencia de múltiples tratamientos. Ocurriría cuando se intentan probar los efectos de dos o más tratamientos simultáneamente. El comportamiento de los sujetos podría ser diferente en presencia o no de sólo uno de esos comportamientos, por lo que si se expone al sujeto a más tratamiento sería difícil generalizar a momentos en que se aplique sólo uno de los tratamientos.

La validez interna y externa a menudo son conceptos relacionados pero inversos, pues cuando mayor control se ejerza sobre un experimento menor la posibilidad de extender los hallazgos a otros contextos más aplicados. Kazdin (1982) identifica nueve dimensiones en las que podría diferir la investigación básica sobre tratamientos, de la aplicación en las situaciones clínicas naturales: (la conducta problema objetivo, la población, la forma de recoger información, los terapeutas, las características de los clientes, la selección del tratamiento, la situación en que se aplica el tratamiento, las variaciones durante el tratamiento, y los procedimientos de evaluación). Cuanto más se parezcan estos factores en la investigación a la situación clínica, mayores garantías de validez externa del experimento.

Sin embargo, debería tenerse en cuenta que unos hallazgos que no estén claros, cuya interpretación resulte confusa, también invalidan su generalización. Una labor prioritaria es asegurar la validez interna de una investigación, primero es asegurar la certeza de los resultados, la imposibilidad de otras explicaciones alternativas, y a partir de ello poder construir progresivamente campos de generalización cada vez más aplicados y clínicos (Kratochwill y Mace, 1984).

Problemas con la generalización clínica

Para el psicólogo clínico la pregunta fundamental debería ser: ¿qué tratamiento, y de qué forma, es más efectivo para este cliente con este grupo de problemas y características?. Una cuestión enfocada dentro de la metodología alrededor del concepto de "validez externa", el grado en que una investigación es generalizable, decir, si el conocimiento que aplicamos en la solución de casos clínicos está basado en investigaciones formales, en experiencias contrastadas informadas por otros, o simplemente en el sentido común.

A menudo la metodología de grupos ha basado el poder de generalización en la cantidad de sujetos utilizados y en su representatividad. Ha desechado la experimentación con un único caso por considerarlo poco "científico" o "cuasi-experimental", por no poder separar la variabilidad individual de los efectos del tratamiento, o por no conseguir validez externa de un estudio clínico.

La estrategia de generalización en los análisis de grupo se basa en la asignación y la selección aleatoria de los sujetos, con la consiguiente inferencia de afirmaciones generales sobre la eficacia de los tratamientos basada en la representatividad de las muestras. Desafortunadamente, siguiendo la argumentación de Hayes (1983) nunca se podrá seleccionar una muestra clínica aleatoria de la población general con problemas, algunos no acuden a tratamiento y los que lo hacen no presentan características aleatorias. Incluso si resolviésemos ese problema, no podríamos forzar a todos los sujetos a participar en una investigación, ni se suelen seguir estrategias completamente aleatorias con ellos. Además, si resolviésemos ese problema, los

resultados de una investigación clínica se podrían generalizar sólo a otras muestras representativas de esa población con características similares a la estudiada, no a otras poblaciones diferentes, ni a muestras con algún sesgo inicial.

La asignación aleatoria no asegura tampoco la generalización hacia abajo, hacia el individuo. La mayor parte de los diseños de grupo intentan detectar qué grupo mejora con el tratamiento frente al "ruido de fondo" que es la variabilidad individual. Aquí, al considerar el caso individual, los errores de medición y los sesgos de variables extrañas son inseparables de los efectos del tratamiento. Se puede afirmar ese efecto respecto al grupo, pero en un caso específico dentro de ese grupo la mejora obtenida no puede afirmarse que haya sido debida exclusivamente al tratamiento. La asignación aleatoria de factores como la maduración, la edad, los errores de medición, etc. iguala su influencia respecto al grupo, no permite las comparaciones individuales (Kazdin, 1981).

Una estrategia para obtener generalización sería correlacionar, una y otra vez, las características de los clientes y sus problemas con los resultados del tratamiento, y obtener así reglas de generalización para los clínicos. Otra solución sería el análisis intensivo del caso individual, múltiples veces, repitiendo la experimentación y tratando de identificar los factores extraños -la variabilidad- y las variables necesarias y suficientes que expliquen los resultados del tratamiento (Kazdin, 1981). De esta forma, se podrían obtener reglas más apropiadas para generalizar a un individuo con unas características determinadas y obtener unos resultados esperados.

La repetición de la experimentación sería la forma de asegurar la generabilidad de unas conclusiones determinadas y de extenderlas a diversos tipos de sujetos, experimentadores, situaciones, etc. La diferenciación ya clásica de Sidman (1973) define: (1) estudios de "**replicación directa**", aquellos en los que se repite el experimento en condiciones idénticas al inicial, puede realizarse repitiendo las mismas condiciones en varias ocasiones en el mismo sujeto o en varios sujetos de idénticas características. (2) estudios de "**replicación sistemática**", en los que se repite la misma experimentación pero cambiando alguna condición de aplicación, bien sea el contexto, alguna de las características del sujeto, el experimentador o terapeuta, la conducta problemática, etc. La repetición de un estudio pero con pequeños cambios en las condiciones de una a otra ocasión sería la mejor forma de obtener una generalización cada vez mayor sobre las conclusiones. Generalmente se realiza con diferentes sujetos y diferentes contextos de experimentación, pero también es admisible en algunos diseños de caso único que permiten cambiar los niveles de algunas de esas variables. Un tipo especial de replicación sistemática sería la "**replicación clínica**", en la que se intenta probar la generalidad de la eficacia de un determinado tratamiento específico, respecto a varios tipos de conductas problemáticas, sujetos, terapeutas, situaciones, etc., en suma, establecer la generabilidad de una terapia.

Barlow y Hersen (1988) definen tres tipos de generalización que serían relevantes al campo clínico: (1) la generalización de los hallazgos entre sujetos o clientes, (2) a través de los agentes de cambio o terapeutas, y (3) a la variedad de entornos en los que se encuentran los clientes.

Hayes (1983, pag. 183) resume en una serie de preguntas clínicas la forma de identificar el poder de generalización de un determinado estudio, una serie de reglas críticas que el clínico debería tener en cuenta:

1. Descripción adecuada: ¿Se han descrito con detalle los pacientes? ¿Tiene mi paciente las mismas características, al menos en sus aspectos más relevantes?.
2. Replicación: ¿Se han descrito los procedimientos de forma detallada que me permitan repetir lo que ellos hicieron?.
3. Adecuación del tratamiento: ¿Comprobaron la integridad del tratamiento, es decir, su correcta aplicación? ¿Hubo observadores o jueces externos al tratamiento?.
4. Condiciones de aplicación: Especificaron las condiciones (terapeutas, ambiente clínico) bajo las que se aplicó el tratamiento? ¿Son similares a las mías? ¿Hubo efectos del terapeuta?
5. Evaluación continua: ¿Se tomaron medidas repetidas a través del tiempo, de forma que se obtuviese una muestra individual adecuada del progreso del tratamiento?.
6. Multimétodo: ¿Hubo diferentes mediciones, si no hay una única medida generalmente aceptada para ese problema?.
7. Factores relevantes: ¿Están relacionadas las características individuales con los resultados del tratamiento? ¿Tiene mi paciente las características más favorables?.
8. Valoración de resultados: Si los resultados informados son respecto al grupo, ¿se informa sobre el porcentaje de individuos que mostraron el efecto en cuestión? ¿Es elevado? ¿Aparecen también datos individuales?.
9. Replicación sistemática: ¿Se han replicado los resultados? ¿Cuántas veces? ¿Por otros investigadores?.
10. Validez clínica: ¿Los efectos -y las diferencias entre efectos- son suficientemente grandes y clínicamente significativos?.
11. Seguimiento y generalización: ¿Ha comprobado el estudio las diferencias que podrían favorecer la generalización? Y si es así, ¿esas condiciones favorables se presentan en mi paciente?.
12. Replicación propia: ¿He probado el procedimiento descrito en otros clientes? ¿Obtuve resultados similares?.
13. Validez interna: ¿Tiene el estudio validez interna? ¿Son poco probables otras explicaciones para los mismos resultados?.

Este tipo de consejos intentan que el clínico produzca y se base en conocimientos que estén consistentemente probados, y que sean congruentes con los principios de una metodología científica y los objetivos de la psicología clínica.

Referencias:

- Barlow, D. H. y Hayes, S.C. (1979). Alternating treatment design: One strategy for comparing the effects of two treatment in a single subject. *Journal of Applied Behavior Analysis*, 12, 199-210.
- Barlow, D. H. y Hersen, M. (1988). *Diseños experimentales de caso único*. Barcelona: Martínez Roca. (Original 1984).
- Campbell, D. T. y Stanley, J.C. (1973). *Diseños experimentales y cuasi-experimentales en la investigación social*. Buenos Aires: Amorrortu. (Original 1963).
- Cook, D. T. y Campbell, D.T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.
- Hayes, S. C. (1983). The role of the individual case in the production and consumption of clinical knowledge. En M. Hersen, A.E. Kazdin y A.S. Bellack (Eds.): *The clinical psychology handbook*. New York: Pergamon Press.
- Johnston, J. M. y Pennypacker, H.S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, N.J.: Lawrence Erlbaum Ass.
- Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology*, 49, 183-192.
- Kazdin, A. E. (1982). *Single-case research designs. Methods for clinical and applied settings*. New York: Oxford University Press.
- Kiesler, D. J. (1981). *Handbook of psychotherapy and behavior change*. New York: John Wiley and Sons.
- Kratochwill, T. R. y Mace, F.C. (1984). Time-series research in psychotherapy. En M. Hersen, L. Michelson y A.S. Bellack (Eds.): *Issues in psychotherapy research*. New York: Plenum Press.
- O'Leary, K. D. y Bokovec, T.D. (1969). Conceptual, methodological, and ethical problems of placebo groups in psychotherapy research. *American Psychologist*, 33, 821-839.
- Rosenthal, R. (1966). *Experimenter bias in behavioral research*. New York: Appleton-Century-Crofts.
- Sidman, M. (1973). *Tácticas de investigación científica*. Barcelona: Fontanella. (Original 1960).
- Ulman, J. D. y Sulzer-Azaroff, B. (1975). Multielement baseline design in educational research. En E. Ramp y G. Semb (Eds.) *Behavior analysis: Areas or research and application*. New Jersey: Prentice-Hall.
- Wilkins, W. (1977). Experiments in applied settings. En A. Gurman y A. Razin (Eds.). *Effective psychotherapy: A handbook of research*. New York: Pergamon Press.