

Diseños de evaluación de programas: bases metodológicas

Salvador Chacón Moscoso, M^a Teresa Anguera Argilaga* y José López Ruiz
Universidad de Sevilla y * Universidad de Barcelona

Las circunstancias inestables que caracterizan el contexto en el que se realiza la evaluación de programas de intervención hacen que sea prácticamente imposible plantear diseños de evaluación estándares. El objetivo de este trabajo es describir una serie de dimensiones estructurales de diseños de evaluación que se puedan implementar de una manera flexible en los contextos de intervención. Desde estos referentes de diseño de evaluación se intentará potenciar la obtención de información con el mayor grado de validez científica, sin arraigar la idea de una lista de diseños que pueden aplicarse ante determinadas situaciones estándares. El esquema de análisis de partida lo estructuramos en tres grandes dimensiones estructurales de diseño: Usuarios del programa, naturaleza de los datos, y momento temporal de registro. A partir de la combinación de estos ejes se justificarán el uso de un tipo de diseños u otros.

Program evaluation designs: methodological issues. Unstable circumstances of program evaluation contexts make impossible a systematization of program evaluation designs. The main objective of this paper is to describe the structural dimensions of program evaluation designs in order to implement them in a flexible way in different evaluation contexts. These dimensions will try to enhance scientific validity in evaluative data, not offering the conception of standard design structures for evaluations. The structural dimensions are: clients of the program, type of data, and recording moments. The combination of these dimensions will justify different evaluative designs.

La evaluación de programas en sus distintos aspectos, sigue las reglas del método científico, aunque la investigación evaluativa no se reduce a éstas, constituyendo hoy un área específica y separada de la anterior. Esta separación y la formación de un «*corpus*» propio, se debe a las peculiaridades de la investigación evaluativa, derivadas en primer lugar de su insistencia en los aspectos de *valor* de los programas (Scriven, 1980); en segundo lugar, la investigación evaluativa se diferencia por su *aplicabilidad*, tanto desde el punto de vista *formativo*, proporcionando *feedback* a los programadores y administradores de los programas (Vedung, 1993, 1996), como en la *toma de decisiones* sobre los resultados de éstos, su continuidad, modificabilidad, etc. (Weiss, 1983; Wholey, 1987; Martínez-Arias, 1996).

Desde este planteamiento entendemos que la evaluación de programas se desarrolla a lo largo de un proceso lógico que sustancialmente no difiere del proceso de investigación en un ámbito aplicado. Se cuenta con una realidad compleja, pero tangible, con programas en que se implementan acciones que a veces no se ajustan al calendario, o que no se ejecutan por igual en todos los sujetos, pero en los cuales el qué, el cómo y el cuándo son registrables. El sector en que se ubica la necesidad, las características del entorno en que se enclava, y la propia naturaleza de la carencia con-

dicionan la amplia casuística de programas, y sobre todo, les imponen fuertes limitaciones que chocan frontalmente con los requisitos que impone el rigor del método científico (Anguera y Chacón, en prensa, a).

De forma ilustrativa, podemos pensar en la *selección de sujetos*. En los manuales metodológicos, y simplificando mucho, se insiste en la práctica —prácticamente se fuerza a su opción— del muestreo probabilístico si se opta por una vía deductiva, y por tanto, siempre que nos situemos bajo la cobertura de un marco teórico consolidado. Este muestreo probabilístico, por ser equiprobable, además de representativo, ¿garantizaría el componente de equidad en los potenciales usuarios? En el momento en que bajamos a la arena de lo cotidiano en evaluación de programas, si tenemos que evaluar un programa domiciliario de atención geriátrica se descarta de entrada una respuesta afirmativa, ya que serán los propios usuarios, o sus allegados, o los responsables de Servicios Sociales de la zona, los que tratarán de recabar la adscripción al programa. Aquí no cabe de plano el muestreo probabilístico (los distintos casos no son equiprobables, ni se eligen al azar, y ni siquiera se puede afirmar, al menos de forma general, que sean representativos del colectivo afectado por la necesidad), independientemente de que exista o no un determinado marco teórico relativo al problema y a su probable intervención, e independientemente también de que existan suficientes recursos (humanos, temporales, económicos, etc.) para que todos ellos sean atendidos. En consecuencia, no podemos hablar de equidad desde el mismo momento en que se plantea una selección mediante muestreo probabilístico de usuarios.

Siguiendo con la selección de sujetos, si se plantea la cuestión desde la vía inductiva de forma obligada o deseada, sea porque no

Correspondencia: Salvador Chacón Moscoso
Facultad de Psicología
Universidad de Sevilla
41005 Sevilla (Spain)
E-mail: schacon@psicoexp.us.es

existe un determinado marco teórico de referencia, o porque no nos interesan los que hayan, deberá seleccionarse un caso único inicial al que le sigue una progresiva acumulación de casos afines, con el fin de llegar a encontrar regularidades en el comportamiento de todos estos casos y que se pueda ir trazando un esquema de funcionamiento del caso general. Yendo a la realidad del profesional, y desde la mayor flexibilidad que permite el planteamiento inductivo, si bien se establecerán criterios para la inclusión en el programa en función de las características técnicas de la necesidad y las condiciones físicas, personales, económicas, familiares, etc., del potencial usuario, es constatable que no siempre se tratará de casos *afines*, al menos en el sentido restrictivo del término. Consecuentemente, tampoco se llevaría a cabo la selección de sujetos de forma metodológicamente correcta, además de seguir cuestionándonos el concepto de equidad en los potenciales usuarios.

Pero podemos pensar en otros tipos de programas, como los institucionales relativos a preservar la capa de ozono o mantener limpia una ciudad o una superficie; o los sanitarios sobre hábitos higiénicos, o la deshabituación al tabaco o a drogas diversas; o los de educación vial. En todos ellos existen componentes que escapan —por la propia complejidad de la realidad— a la norma científica, lo cual da lugar a dos alternativas por las que entendemos que es fácil optar: se flexibiliza la norma, o se está al margen de toda norma.

La segunda posibilidad dejaría a la evaluación de programas fuera de un ámbito formal de estudio por disciplinas como la Psicología, Sociología, Medicina, Educación, etc. Se podrían describir *casos* aislados, como se ha hecho desde determinadas corrientes radicales de metodología cualitativa —ver, por ejemplo, Smith y Cantley (1985), o el número monográfico de *Qualitative Health Research* editado por Engel (1992), o el de *Qualitative Inquiry* editado por Reason y Lincoln (1996)—, pero es indudable que masivamente interesa ajustarnos a la lógica del procedimiento científico.

Esta lógica, sin embargo, que responde a los principios del positivismo científico, debe ser sensible a las específicas características individuales, situacionales, del programa, de los recursos disponibles, etc. La casuística es amplísima, y no es fácil lograr el equilibrio entre el rigor imprescindible y la flexibilidad adaptativa respecto al programa a evaluar.

Todas estas circunstancias inestables que caracterizan el contexto en el que se realizan las evaluaciones de programa de intervención hacen que sea prácticamente imposible plantear estructuras estándares de diseño apriorísticas. En este sentido el objetivo de este trabajo es describir una serie de dimensiones estructurales de diseño que puedan implementarse de una manera flexible en los contextos de intervención particulares. Desde estos referentes de diseño de evaluación se intentará potenciar la obtención de información con el mayor grado de validez científica, sin arraigar la idea de una serie de estructuras de diseño rígidas que sólo pueden aplicarse en situaciones estándares (Shadish, Cook y Campbell, en preparación).

Dimensiones estructurales de diseños de evaluación

En la literatura se proponen distintas formas de clasificar los diseños que podrían aplicarse a la evaluación de programas (podemos referirnos a algunos autores, como Arnau, Anguera y Gómez, 1990; Ato, 1991; Judd y Kenny, 1981; Kazdin, 1980; Kish, 1987; Owen y Rogers, 1999). A pesar de ello resulta difícil encajar estas clasificaciones en la práctica profesional de la evaluación de pro-

gramas. Esta situación se da fundamentalmente porque en el contexto de la evaluación de las intervenciones es complejo encontrar modelos o referentes teóricos a partir de los cuales disponer de criterios con los que desarrollar un modelo de selección de los distintos componentes que tendría que analizar un programa de evaluación. Tengamos presente que no sólo nos referimos a la selección de usuarios del programa a evaluar, tal y como planteábamos en el ejemplo del apartado anterior; es decir, a qué población de sujetos nos referimos exactamente; también hemos de tener en cuenta cuáles son los distintos tipos de intervenciones teóricas que podrían utilizarse en cada casuística, cuáles son las variables que vamos a medir, cuál es la delimitación realizada del contexto de intervención, en qué momento o momentos concretos y con qué recursos vamos a registrar los datos... A estos interrogantes, que aunque aparentemente simples son de hondo calado, se les une el problema de conseguir una alta fiabilidad en las medidas (por ejemplo, para evaluar el impacto de un programa).

Todas estas circunstancias hacen que nos estemos refiriendo a un ámbito de trabajo en el que impera una filosofía falsacionista, ya que el grado de conocimiento previo y de control sobre el objeto de evaluación es reducido. En este sentido se habrán de prever cuáles pueden ser las principales amenazas a la validez, tanto de representatividad como de control, que puede presentar nuestra evaluación particular. El problema de esta situación es la dificultad de disponer de teorías sobre las posibles interacciones entre las variables en los contextos de intervención particular; más aún cuando unas mismas variables pueden interactuar de manera distinta en un mismo contexto en momentos temporales distintos.

En el marco de esta realidad descrita abogamos por justificar el desarrollo de un diseño de evaluación desde unos referentes dimensionales que se combinarán de distintas formas dependiendo del contexto de intervención particular donde se implementen. El esquema de partida lo estructuramos en tres grandes dimensiones: *usuarios del programa*, *naturaleza de los datos*, y *momento temporal* (Anguera y Chacón, en prensa, b). Para establecerlas hemos partido de una idea original de Cattell (1952), que ya a mitad de siglo empezó a producir sus frutos. Utilizaba un paralelepípedo en el que las aristas representaban personas, variables de medida y ocasiones de medida para poder ilustrar los datos en estudios de covariación. De esta forma, cada cara del paralelepípedo permite obtener una matriz bidimensional de puntuaciones, ya que la definen dos aristas, y la tercera cara no se muestra, pero se fija. Muchos años después, Nesselroade y Hershberger (1993), en sus estudios sobre población, la adaptan para explicar la variabilidad intraindividual (Nesselroade, 1988, 1991) utilizando las mismas dimensiones de personas, variables de medida y ocasiones de medida.

En evaluación de programas consideramos que existen tres referentes desde una perspectiva metodológica: A) a quiénes va dirigido el programa, ya que de lo contrario éste perdería su razón de ser, motivo por el que los *usuarios* ocupan el primer lugar. B) tipo de información que se obtiene, habitualmente de carácter cambiante a lo largo del proceso de implementación y en función de las diversas acciones que se llevan a cabo, por lo que la *naturaleza de los datos* es un referente obligado. C) El carácter sincrónico o diacrónico del proceso de evaluación, dependiendo de si nos situamos en la evaluación sumativa o formativa, respectivamente (siguiendo con los presupuestos del modelo lineal vs. no lineal de Veney y Kaluzny —1984).

Configuración de diseños de evaluación

De acuerdo con el complejo marco en el que se desarrolla un programa de evaluación hemos justificado la dificultad de poder establecer una relación de diseños estándares de evaluación. Esta situación justifica la necesidad de configurar los diseños evaluativos de acuerdo con las necesidades y características del programa concreto, y por tanto desde una combinación de las dimensiones de diseño anteriormente señaladas. El esquema básico resultante de la consideración conjunta de todos los referentes comentados lo podemos configurar a partir del cruce de las dimensiones de usuarios y temporalidad, en combinación con el grado de intervención o dominio que se tenga sobre el contexto de evaluación (Fig. 1):

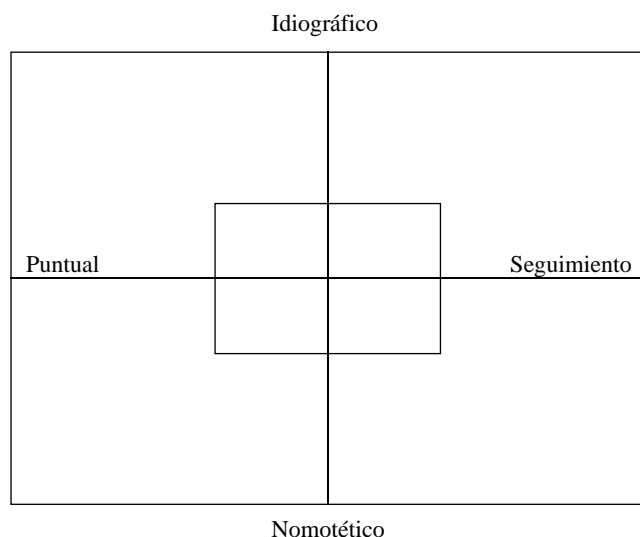


Figura 1. Esquema básico para la configuración de diseños de evaluación de programas (Anguera, 1995b)

A. Carácter ideográfico vs. Nomotético (eje vertical)

El eje vertical se refiere al carácter ideográfico o nomotético en función de los usuarios del programa de intervención. No siempre adquiere la misma relevancia, pero plantea importantes cuestiones a nivel metodológico (Posavac y Carey, 1985) según se trate de sujetos individualmente considerados o de una colectividad (o muestra representativa de ella), si atendemos a la propuesta clásica de Allport (1942) en relación a los términos *ideográfico vs. no - motético*.

Ahora bien, a tal propuesta se han incorporado variantes adaptativas a las diversas situaciones evaluativas:

a) Se considerarán también como ideográficos estudios que amplían o restringen la propuesta clásica consistente en un individuo. Por una parte, entre los primeros se hallarán todos aquellos casos en que los usuarios son varios individuos entre los cuales existe un criterio de afinidad, agrupación, o reglas del juego a seguir; por ejemplo, un programa de intervención familiar, independientemente de cuántas personas componen aquella unidad familiar y de cuantos niveles de respuesta se registren. Y, por otra parte, los que restringen el concepto clásico de ideográfico se centran en un solo nivel de respuesta, sea de un individuo único, o de varios; por ejemplo, si consideramos únicamente el nivel de conducta verbal, y evaluamos la resolución de un conflicto entre

miembros de una familia a partir de la discusión de los respectivos puntos de vista y balance entre pros y contras de cada opción de solución.

b) Nomotéticas serán también aquellas variantes en que, independientemente de que tengamos un usuario o un grupo de usuarios, interesan varios niveles de respuesta. Así, en un programa de atención psicológica a enfermos infartados y sus familiares interesan los niveles de respuesta verbal y no verbal. Luego, nomotéticas serán todos aquellos diseños evaluativos en que se configura un elemento de pluralidad de unidades, sean individuos (propuesta clásica) o niveles de respuesta (variante posteriormente introducida).

Una vez delimitado el número de participantes en un estudio evaluativo (individuos sobre los que se interviene), el evaluador deberá decidir si todos deberán formar parte o no de la correspondiente evaluación, o incluso vincular esta decisión a distintas fases del proceso.

Los principales argumentos a favor de la inclusión del colectivo completo o de una muestra representativa del mismo son de índole metodológica (análisis de los efectos en toda la cobertura relativa a personal), pero también ética, y, en ocasiones, política. Así, si un centro hospitalario ofrece una unidad de atención psicológica a enfermos infartados por segunda vez, y es mayor la demanda que el número de plazas que puede cubrir el servicio, ¿se podría hablar de un criterio más «ético» que otros?, ¿cabe, desde los principios éticos, extraer una muestra representativa cuando todo el colectivo presenta un mismo tipo de necesidad demandada?

No faltan tampoco argumentos para el estudio evaluativo de los efectos de un programa en fases diferenciadas, de forma que a un primer análisis efectuado de forma nomotética le sigue un segundo basado en el estudio de sujetos individuales. La perspectiva ideográfica, desde la expansión e incidencia actual de la Psicología de las diferencias individuales, está alcanzando una gran relevancia en la implementación y evaluación de programas sociales y sanitarios. Cada vez los profesionales son más sensibles a la consideración diferencial de sujetos que por su trayectoria vivida (circunstancias personales, «event-life», rasgos de personalidad, etc.), requieren un análisis específico e individualizado de los efectos de un determinado programa de intervención. Pensemos en niños con trastornos comportamentales en el aula y con historias de vida absolutamente distintas (Herrero, 1989), o en enfermos con repetidos infartos que tenían muy diferente nivel de calidad de vida (Tuset, 1990), o en internos penitenciarios que cumplen una condena de igual duración a partir de un historial personal y delictivo completamente distinto (Redondo, 1992), o en deportistas que presentan determinadas peculiaridades en sus tácticas de juego (Hernández Mendo y Anguera, 1998).

B. Temporalidad del registro (eje horizontal)

La configuración básica de los diseños permite distinguir entre *registro puntual y seguimiento*. El registro puntual permitirá realizar un análisis de la situación en un momento dado en el tiempo, mientras que el seguimiento implica disponer de un determinado número de sesiones a lo largo del período de implementación del programa.

El criterio de temporalidad en el registro permite tener también en cuenta el punto de partida (previo, durante, o después de la apli-

cación de la intervención, o, expresado en otros términos, de la implementación del programa), y el periodo de cobertura en la recogida de datos (hasta el fin de la intervención, seguimientos puntuales periódicos hasta un determinado momento, idem con un seguimiento continuo, etc.).

Es muy fácil de argumentar cuál es el óptimo o ideal, partiendo del presupuesto de la existencia de recursos suficientes. Evidentemente, desde antes del inicio de la intervención, durante el tiempo que implique su puesta en práctica, y efectuando un seguimiento posterior a medio o largo plazo que posibilite un análisis riguroso de los efectos del programa.

Ahora bien, las distintas posibilidades que implica el barajar estos elementos, la necesidad de adecuarse a recursos generalmente limitados (Fienberg y Tanur, 1987), y la propia naturaleza de la intervención, deben dar lugar a las decisiones relativas al registro (cómo, desde cuándo, hasta cuándo, con qué periodicidad, con qué garantías en la formación del personal que participa en la evaluación, etc.), en el más amplio sentido del término (Blanco y Anguera, 1991).

Finalmente, será conveniente distinguir, en este criterio relativo al carácter continuo o discreto del registro a lo largo del tiempo, entre la recogida de datos actuales (Plewis, 1985), y los retrospectivos (Holland y Rubin, 1988), como por ejemplo los referidos a material de archivo, tanto si se trata de datos censales o estadísticos, como de protocolos o informes personales (autobiográficos o realizados por terceras personas), siempre que se mantenga la homogeneidad de los criterios seguidos en su recogida y, en su caso, codificación (como categorías en un análisis de contenido de autoinformes), y no entorpezca su utilización el tan frecuente problema de los «missing data» (Little y Rubin, 1987).

C. Grado de intervención o dominio sobre el contexto de evaluación (recuadro interior vs. exterior del recuadro).

A estas dos dimensiones se ha de incorporar el referente del grado de intervención sobre la situación. Hemos intentado simbolizar esta dimensión presentando un recuadro interior en los ejes de la figura 1. Se pretende representar una superposición de dos planos distintos, de mayor a menor intervención respectivamente, dependiendo de si se está más hacia el centro o hacia el exterior de dicho recuadro interior. Con el término intervención en la situación hacemos referencia a la naturalidad de la situación, es decir el grado en que la relación de los sujetos, usuarios del programa, modifican sus interacciones naturales con el medio. Por ejemplo, en un programa de natación en la tercera edad en que los participantes realizan habitualmente ésta u otras actividades deportivas nos estaríamos refiriendo a un diseño de evaluación de baja intervención (este programa estaría representado en la parte exterior del recuadro). En cambio si nos interesase un diseño en el que un grupo de familias marginales de una gran ciudad han sido seleccionadas al azar y asignadas posteriormente a pueblos con bajo número de habitantes de otras provincias, para con ello disminuir la probabilidad de que los hijos de estas familias realicen conductas delictivas, se trataría de un tipo de programa con un alto grado de intervención (este programa estaría representado más hacia el interior del recuadro central de la figura 1).

El concepto de intervención no es dicotómico, no podemos establecer que haya o no intervención, se trata de un concepto de

grado. De hecho no tendría sentido plantear la no existencia de intervención en tanto que el propio hecho de desarrollar una evaluación (con sistemas de registro, instrucciones a los usuarios en su caso,...) supone una intervención en sí misma. Los niveles de intervención se pueden desglosar de diversas formas. La opción que defendemos en este trabajo distingue entre diseños de intervención baja y diseños de intervención media-alta (parte exterior o interior, respectivamente, en el recuadro central de la figura 1). Se opta por esta dicotomización de la gradación en el nivel de intervención porque parece claro cuáles pueden ser los extremos de este eje bipolar, las denominadas metodologías naturalistas y experimentales, pero una vez que nos adentramos en ese teórico continuo es difícil establecer criterios que nos delimiten los límites a partir de los cuales referimos a un tipo de metodología u otra. Más aún cuando en el ámbito de la intervención real suelen ser varios los procedimientos utilizados en un mismo programa de evaluación.

Desde esta argumentación, y obviando que en la práctica evaluativa se suelen utilizar conjuntamente distintos tipos de procedimientos, de una forma simplificada incluiríamos en los denominados diseños evaluativos de intervención baja a los diseños observacionales (diacrónicos, sincrónicos y mixtos), e incluiríamos en los diseños evaluativos de intervención media-alta a los diseños de grupo control no equivalente, de discontinuidad en la regresión y de series temporales interrumpidas.

A pesar de que hemos acabado desembocando en una interesante discusión, que sin duda requiere una mayor extensión en su tratamiento, consideramos que la cuestión no radica tanto en poder delimitar con precisión el grado en que se está interviniendo, sino más bien en tenerlo presente y estudiado en lo posible. Con ello se intentará conocer que posibles fuentes de variación estamos incorporando en los datos registrados. A partir de este análisis detallado podremos disponer de más elementos de juicio a partir de los cuales valorar el grado de validez de la información registrada, y a partir de la cual se van a tomar decisiones que en la mayoría de los casos tienen una importante repercusión social.

A modo de reflexión final

En este trabajo se ha querido enfatizar que para el diseño de la evaluación de un programa nos basamos más en un criterio de «plausibilidad» del conocimiento de la mayor parte posible de variables que puedan implicar amenazas a la validez de la información. De ahí que los elementos de diseños planteados se basen por una parte en estudiar que posibles amenazas de validez existen en el contexto de evaluación, y por otra en intentar neutralizar o al menos minimizar sus efectos en la mayor medida posible. Esta circunstancia hace que la calidad de los diseños evaluativos dependa de los estudios previos existentes y de la sistematización de las posibles amenazas a la validez con las que nos podemos encontrar en los distintos ámbitos de intervención. En la mayor parte de los diseños de evaluación no existe un mecanismo ómnibus como la asignación aleatoria que pueda neutralizar las amenazas a la validez de una manera directa. Por tanto los controles se han de basar en los recursos de los diseños planificados directamente ligados a la robustez de las teorías y medidas utilizadas.

Referencias

- Allport, G.W. (1942). *The use of personal documents in psychological science*. Nueva York: Social Science Research.
- Anguera, M.T. y Chacón, S. (en prensa, a). Bases Metodológicas. En Anguera Argilaga, M.T. (Dir.). *Evaluación de programas sociales y sanitarios: Un abordaje metodológico*. Madrid: Síntesis.
- Anguera, M.T. y Chacón, S. (en prensa, b). Dimensiones estructurales de diseño para la evaluación de programas. *Apuntes de Psicología*.
- Arnau, J.; Anguera, M.T. y Gómez, J. (1990). *Metodología de la investigación en ciencias del comportamiento*. Murcia: Servicio de publicaciones de la Universidad de Murcia.
- Ato, M. (1991). *Investigación en ciencias del comportamiento. I: Fundamentos*. Barcelona: P.P.U.
- Blanco, A. y Anguera, M.T. (1991). Sistemas de codificación. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica* (pp. 193-239). Barcelona: P.P.U., Vol. I.
- Cattell, R.B. (1952). The three basic factor analytic research designs – their interrelationships and derivatives. *Psychological Bulletin*, 49, 499-520.
- Engel, J.D. (Ed.) (1992). Issues of methodology in qualitative inquiry [monographic number]. *Qualitative Health Research*, 2 (4), número completo.
- Fienberg, S.B. y Tanur, J. (1987). The design and analysis of longitudinal surveys: Controversies and issues of costs and continuity. En R.F. Boruch y R.W. Pearson (Eds.), *Designing research with scarce resources*. Nueva York: Springer-Verlag.
- Hernández Mendo, A. y Anguera, M.T. (1998). Análisis de coordenadas polares en el estudio de las diferencias individuales de la acción de juego. En M.P. Sánchez y M.A. Quiroga (Coords.), *Perspectivas actuales en la investigación psicológica de las diferencias individuales* (pp. 84-88). Madrid: Centro de Estudios Ramón Areces.
- Herrero, M.L. (1989). *Incidencias de la historia personal en el comportamiento en el aula: Estudio observacional analítico*. Tesis Doctoral no publicada. Barcelona: Universidad de Barcelona.
- Holland, P.W. y Rubin, D.B. (1988). Causal inference in retrospective studies. *Evaluation Review*, 12 (3), 203-231.
- Judd, C. y Kenny, D. (1981). *Estimating the impact of social interventions*. Cambridge M.A.: Cambridge University Press.
- Kazdin, A. E. (1980). *Research design in clinical psychology*. Nueva York: Harper y Row.
- Kish, L. (1987). *Survey sampling*. Nueva York: John Wiley
- Little, R.J. y Rubin, D.B. (1987). *Statistical analysis with missing data*. Nueva York: Wiley
- Nesselroade, J.R. (1988). Sampling and generalizability: Adult development and aging research issues examined within the general methodological framework of selection. In K.W. Schaie, R.T. Campbell, W. Meredith & S.C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 13-42). New York: Springer.
- Nesselroade, J.R. (1991). Interindividual differences in intraindividual changes. In J.L. Horn & L. Collins (Eds.), *Best methods for measuring change*. Washington: American Psychological Association.
- Nesselroade, J.R. y Hershberger, S.L. (1993). Intraindividual variability: Methodological issues for population health research. In K. Dean (Ed.), *Population health research. Linking theory and methods* (pp. 74-94). Londres: Sage.
- Owen, J.M. y Rogers, P.J. (1999). *Program evaluation. Forms and approaches*. Londres: Sage.
- Plewis, I. (1985). *Analysing change. Measurement and explanation using longitudinal data*. Nueva York: Wiley.
- Posavac, E.J. y Carey, R.G. (1985). *Program evaluation. Methods and case studies*. Nueva York: Prentice-Hall.
- Reason, P. y Lincoln, Y.S. (Eds.) (1996). Quality in Human Inquiry (Special issue). *Qualitative Inquiry*, 2 (1), whole number.
- Redondo, S. (1992). *Evaluar e intervenir en prisiones*. Barcelona: P.P.U.
- Smith, G. & Cantley, C. (1985). Policy evaluation: The use of varied data in a study of a psychogeriatric service. In R. Walker (Ed.), *Applied qualitative research* (pp. 156-174). Aldershot: Gower.
- Tuset, A. (1990). *Análisis de las respuestas al test de Rorschach de un grupo de sujetos afectados de un primer infarto de miocardio*. Tesis Doctoral no publicada. Barcelona: Universidad de Barcelona.
- Veney, J.E. y Kaluzny, A.D. (1984). *Evaluation and decision making for health services program*. Englewood Cliff, N.J: Prentice Hall.