

## La evaluación de la eficiencia en la intervención familiar: generalizabilidad y optimización del *Programa Experiencial para Padres*

M<sup>a</sup> Luisa Máiquez Chaves, Angel Blanco-Villaseñor\*, M<sup>a</sup> José Rodrigo López e Ignace P. R. Vermaes\*\*  
Universidad de La Laguna, \* Universidad de Barcelona, \*\* Catholic University of Nijmegen (Países Bajos)

Cada día se hace más necesaria la evaluación tanto de la eficacia como de la eficiencia de los programas de intervención familiar. El objetivo de este estudio es presentar el modelo de evaluación de la eficiencia seguido con el *Programa Experiencial para Padres* (Máiquez, Rodrigo, Capote y Vermaes, 2000). Este programa se desarrolló en tres centros comunitarios de la isla de Tenerife, destinados a prevención primaria y secundaria. El Programa consta de 4 Módulos repartidos en un total de 12 reuniones. En cada una de ellas se realizaban observaciones al grupo de padres y al guía-mediador, llevando así un seguimiento del proceso. La teoría de la generalizabilidad nos permite comprobar la validez de la estructura del diseño que hemos utilizado para poder generalizar los resultados. Encontramos que se puede mejorar la aplicabilidad del programa incrementando moderadamente el número de categorías y disminuyendo el número de observaciones y registros. De esta forma se reduce el coste del programa en personal observador sin perder información.

*Efficiency estimation model based on the Experiential Program for Parents: A Generalizability Study.*  
There is a growing concern on the estimation of both efficacy and efficiency of parenting intervention programs. The present study is aimed at showing the efficiency estimation model based on the *Experiential Program for Parents* (Máiquez, Rodrigo, Capote y Vermaes, 2000). This Program was developed in three different community centers placed on Tenerife island, which were dedicated to primary and secondary prevention. The Program is made up of 4 Modules to be treated in 12 meetings. In all these meeting parents and the assistant-mediator were observed to have a follow-up of the process of knowledge building within the group. Generalizability theory allowed us to prove the reliability of the design structure used in order to generalize the results. We found that a improvement of the Program application is possible, on the one hand, by increasing moderately the number of categories and, on the other hand, by diminishing the number of observations and registers. In this sense, we reduce Program costs concerning assistant staff while keeping the information level about the progress observed.

En los últimos años se han desarrollado multitud de programas de intervención familiar basados en modelos teóricos diferentes, orientados a objetivos muy diversos, dirigidos a poblaciones con situaciones familiares muy variadas, que utilizan un buen número de estrategias metodológicas, con estructuras organizativas muy diversas y que se apoyan en diferentes recursos sociales (ver revisiones de Arruabarrena y de Paúl, 1995; Gracia, 1997; Máiquez, 1997; Vila, 1998). Sin embargo, son escasas las ocasiones en las que se cuenta con una evaluación suficiente y contrastada de la idoneidad de los mismos. Esta es una grave carencia ya que el reto de la intervención familiar se centra en la optimización de los recursos para asignarlos a aquellos programas que hayan demostrado tener un mayor valor y utilidad para la población a la que van destinados. Este artículo se ocupa exclusivamente de la evaluación de los programas grupales de educación para padres desarrollados

en centros comunitarios, uno de los posibles programas de apoyo a la familia que se inscriben preferentemente en el ámbito de la prevención primaria y secundaria (Gracia, 1997). En nuestra opinión, la evaluación de estos programas comprende dos facetas evaluativas: la eficacia y la eficiencia.

Cuando se evalúa la eficacia de un programa (grado por el que un sistema social consigue sus objetivos) se analiza si éste ha producido los cambios esperados en los destinatarios elegidos en relación con los objetivos, la metodología y los contenidos propuestos. Entendida en estos términos la evaluación permite conocer el valor o el mérito del programa según sus propios criterios internos de idoneidad, definidos a partir de los indicadores de cambio propuestos. Además, como estos indicadores se suelen establecer a partir de un modelo teórico, se pueden postular las razones conceptuales del por qué se han producido tales cambios, siempre según dicho modelo.

Cuando se evalúa la eficiencia de un programa el objetivo es averiguar si podemos generalizar los resultados del programa y optimizar algunos de sus parámetros estructurales para mejorar su aplicabilidad en futuras ocasiones. Así, se puede realizar un análisis de los recursos utilizados y su estructuración, por ejemplo, los grupos de padres que se han beneficiado del mismo, la distribución

---

Correspondencia: M<sup>a</sup> Luisa Máiquez Chaves  
Centro Superior de Educación  
Universidad de La Laguna  
38204 Tenerife (Spain)  
E-mail: mmaicha@ull.es

de las sesiones, el número de observaciones realizadas, etc. A partir de esta evaluación se puede conocer la utilidad social del programa en función de los recursos humanos y materiales invertidos en el mismo. Además, se puede mejorar el diseño del programa para que se adapte mejor a la red de servicios sociales en la que planea insertar como un recurso más de dicha red. Por último, pueden realizarse estudios económicos en términos de la relación coste-beneficio del programa (Anguera, 1992, 1995; Rossi & Freeman, 1989), lo cual es muy útil cuando se busca su implantación a gran escala.

Ambas facetas de la evaluación de programas pueden darse indistintamente aunque lo ideal es que sean complementarias. Un programa puede ser eficaz pero no resultar eficiente y viceversa. Piénsese por ejemplo en aquellos programas de intervención que han demostrado ser eficaces en unas condiciones cuasi-experimentales que son muy difíciles de encontrar en la vida real, o bien en aquellos programas con vocación comunitaria pero difíciles de integrar en la red de servicios sociales por suponer demandas de recursos humanos y materiales no disponibles. Alternativamente, piénsese en programas que cumplen estos últimos requisitos de eficiencia y aplicabilidad en la red social, pero cuyo modelo teórico de partida y, por ende, sus indicadores de cambio no son capaces de captar cambios significativos e interesantes en los destinatarios según su problemática familiar. Lo ideal es que un programa pueda superar ambos filtros evaluativos antes de finalizar su puesta a punto para ser ofrecido como un recurso social comunitario.

Este doble proceso evaluativo se ha llevado a cabo con el *Programa Experiencial para Padres* (madres y padres) realizado en la Comunidad Canaria (Máiquez, 1997). En este programa se pretende que los padres reflexionen sobre las teorías implícitas que mantienen respecto al desarrollo y la educación de sus hijos. A través del análisis de las situaciones cotidianas se consigue que los padres sean conscientes de lo que hacen y además conozcan otras formas alternativas de proceder. Este aprendizaje experiencial debe producirse en un escenario sociocultural de construcción compartida y negociada de realidades, en el que se genere un proceso de participación guiada que vaya introduciendo cambios en cogniciones concretas y en la acción situada. La meta es la adquisición de conocimiento aplicable a la vida cotidiana y flexible según las personas y situaciones. La descripción pormenorizada de su modelo teórico, los contenidos del programa así como los detalles de la evaluación de la eficacia del programa están en Máiquez, Rodrigo, Capote y Vermaes (2000). Baste señalar aquí que, como recomiendan Arruabarrena y de Paúl (1995), Fernández-Ballesteros (1995) y Anguera (1990, 1996), para evaluar la eficacia del programa no nos hemos limitado a una evaluación de resultados o de producto, sino que hemos realizado una evaluación del contexto (necesidades y cobertura del programa), una evaluación inicial (perfil sociodemográfico y psicológico de los grupos de padres) y una evaluación bastante exhaustiva del proceso (dinámica de las sesiones a través del comportamiento de los padres y del mediador por medio de unos indicadores de cambio). Así mismo, hemos partido de múltiples registros ya que se han utilizado una gran variedad de indicadores medidos con cuestionarios, escalas de auto-observación y escalas de calificación, cumplimentadas a partir de dos fuentes de información, los observadores y los padres. Por último, otra característica de la evaluación de la eficacia del programa experiencial es el uso combinado de varios tipos de diseño (Cook, 1985, citado por Fernández-Ballesteros, 1995): diseño correlacio-

nal para la evaluación inicial, diseño observacional diacrónico (no-experimental) para la evaluación del proceso y un diseño cuasi-experimental solo-postest grupo control para la evaluación del producto (Campbell & Stanley, 1966; Cook & Campbell, 1979).

El objetivo de este estudio es presentar el modelo de evaluación de la eficiencia seguido con el *Programa Experiencial para Padres*. Partimos de la idea de que la primera aplicación de un programa de intervención es una buena ocasión para replantearse muchas decisiones que se han tomado con objeto de mejorar su eficiencia en futuras aplicaciones del mismo. Por ejemplo, podemos plantearnos mejoras relativas a las observaciones que se realizan en la evaluación del proceso: ¿cuántas categorías serían las apropiadas y las que se pueden observar en cada sesión? o ¿es necesario observar en todas las sesiones y registrar en los tres momentos de la sesión?. En cualquier caso, queremos optimizar el programa manteniendo siempre la validez.

En la búsqueda de generalizar los resultados de la investigación evaluativa, Cronbach asume que la generalizabilidad del conocimiento facilita su transferencia. Podemos definir la generalizabilidad como «el grado por el cual podemos generalizar un resultado obtenido en unas condiciones particulares a un valor teórico buscado» (Blanco-Villaseñor, 1993, p. 189). La validez externa de un programa implica que tales resultados pueden ser generalizados a otros universos de generalización. Esta medida se obtiene a partir del coeficiente de generalizabilidad, que trata de estimar en qué medida la media observada en determinadas condiciones puede ser generalizada a la media de todas las observaciones posibles. Con este estudio pretendemos comprobar la validez de la estructura del diseño que hemos utilizado para poder generalizar los resultados, es decir, si el número de grupos, módulos, sesiones o reuniones, categorías de observación y momentos de registro en cada sesión han sido suficientes.

En concreto los objetivos específicos del estudio serían: Optimizar la calidad y la precisión de la recogida de información *a priori* para generar una estructura de diseño del programa *a posteriori* con el fin de implementar dicho programa en poblaciones más amplias, centrándonos en dos fuentes de recogida de información -la observación al grupo de padres y a los mediadores-. Simultáneamente, minimizar los errores de muestreo generados por la utilización de facetas no-aleatorias en los diseños cuasi-experimentales.

#### Fases de la Teoría de la Generalizabilidad (TG)

La *Teoría de la Generalizabilidad* (TG), que vamos a presentar, ha sido concebida justamente por sus autores (Cronbach, Rajaratnam & Gleser, 1963) para definir el concepto estadístico de muestreo de fuentes de variación múltiples. Cronbach, Gleser, Nanda & Rajaratnam (1972) han podido tratar cada característica de la situación de evaluación (por ejemplo, personalidad del evaluador, estado subjetivo del sujeto objeto de evaluación, característica de la categoría de registro, ocasiones de registro, etc.) como una faceta de un diseño de evaluación. Aplicando las técnicas del análisis de la varianza, podremos cuantificar la importancia de cada fuente de variación.

Es posible entonces definir la puntuación verdadera como la esperanza matemática de todas las observaciones posibles, y el error como una fluctuación muestral correspondiente a la extracción aleatoria de ciertos niveles de las facetas consideradas (elección de determinados evaluadores, de determinados momentos, de dife-

rentes categorías, de diversos grupos...). La teoría estadística puede decirnos el intervalo en que se encontrará la puntuación verdadera cuando utilizemos un tipo de muestreo y qué progresos conseguiremos si seleccionamos de otra forma las muestras.

La TG nos ofrece así un marco más satisfactorio para buscar las estimaciones de precisión de generalización y de márgenes de error, ya que es suficientemente globalizadora como para adaptarse a las condiciones particulares de cada objeto de medida (Blanco-Villaseñor, 1992; Blanco-Villaseñor, Losada & Anguera, 1991). Este carácter globalizador, sin embargo, no ha sido puesto de manifiesto por los autores de esta teoría, quienes la han formulado casi exclusivamente en términos apropiados a contextos psicométricos (García Cueto, 1996), en su obra de base de 1972, *The Dependability of Behavioral Measurement: Theory of Generalizability for Scores and Profiles* (Cronbach, Gleser, Nanda & Rajaratnam, 1972). Las extensiones sucesivas de este enfoque ha planteado nuevas categorías de problemas, que han sido propuestos y resueltos por Berk (1979), Cardinet & Tourneur (1985), Mitchell (1979), Plewis (1988) y Smith & Teeter (1982) en el ámbito de la observación sistemática del comportamiento. A su vez, y como consecuencia de estas extensiones, se ha multiplicado el espectro de aplicaciones diversas en el ámbito de la Evaluación de Programas, como lo demuestran los trabajos recientes en el área de la evaluación de actividades académicas en temas educativos (Cronbach, Linn, Brennan & Hartel, 1997; DeShon, Ployhart & Sacco, 1998; Mcbee & Barnes, 1998; Rindermann, 1997; Sun, Valiga & Gao, 1997), en el área de la evaluación de la actividad física y del deporte (Coleman & Epstein, 1998; Crocker, Bailey, Faulkner, Kowalski & Mcgrath, 1997; Murawski & Miederhoff, 1998; Nattiger, McAuliffe & Schapira, 1997), en el área de la evaluación clínica (Gerin, Christenfeld, Pieper, Derafael, Su, Stroessner, Deich & Pickering, 1998; Tome & Isaac, 1997) y en la evaluación de las organizaciones (Irving, Coleman & Cooper, 1997; Marcoulides, 1997). Todos estos trabajos han contado con literatura científica reciente que solventaba problemas metodológicos de diversa índole (Algina & Keselman, 1997; Cardinet, 1997) y de forma específica en su aplicación a datos categóricos en el trabajo de Li & Lautenschlager (1997).

En cualquier caso y apoyándose en el hecho de que ninguno de los factores incluidos en un diseño de evaluación tiene una primacía particular y que, por ejemplo, la *diferenciación* de las dificultades de los evaluadores podría ser en ciertos casos tan importantes como las de los individuos, estos autores han podido abordar una serie de casos particulares, no considerados por Cronbach et al. (1972), que requieren un nuevo marco conceptual y procedimientos más generales de cálculo.

Las nociones clásicas de puntuación verdadera y del error van a ser reemplazadas por conceptos más en armonía con desarrollos estadísticos modernos. Sea  $X_{ie}$  el valor observado atribuido a cualquiera de los individuos (i) por cualquiera de los evaluadores (e) en una sesión de evaluación y en una competencia particular. Cada una de las *celdas* de la matriz individuos por evaluadores (I x E) contiene, por tanto, diferentes valores  $X$  con subíndices que describen las posibles combinaciones singulares evaluador-individuo.

Cada valor  $X_{ie}$  puede describirse como una suma en la que están implicados tres parámetros  $-\mu_i$ ,  $\mu_e$  y  $\mu$ . La *puntuación universo*, definida en esta situación peculiar como  $\mu_i$ , corresponde al valor medio obtenido por un individuo cualquiera en el conjunto de valores dados por la totalidad de evaluadores del *universo*. Este valor caracteriza la realización de un individuo, en oposición al va-

lor que obtendría en una prueba psicológica a través de una serie de ítemes, que sólo reflejan una muestra de los ítemes del *universo*. El subíndice (e) ha sido eliminado, dado que  $\mu_i$  es el valor medio obtenido por el individuo (i) a través de todos los evaluadores (e).

Formalmente, definimos la *puntuación universo*  $\mu_i$  como el valor esperado (E) de la variable o faceta aleatoria  $X_{ie}$  a través de todos los evaluadores: El valor esperado (E) de una variable aleatoria tal como  $X_{ie}$  es, singularmente, el valor medio obtenido en infinitud de ocasiones. Éste constituye un valor medio obtenido en un número  $k$  de evaluadores (e) que se aproxima al infinito (el límite de  $k$  tiende a infinito,  $k \rightarrow \infty$ ).

El segundo parámetro,  $\mu_e$ , es el valor medio de la *población* obtenido por un evaluador (e). Se define como el valor esperado (E) de  $X_{ie}$  a través de todos los individuos. El tercer parámetro,  $\mu$ , es la media de ambos, *población* y *universo*. Es decir, el valor medio total obtenido a través de la totalidad de la *población* y del *universo*.

Los parámetros  $\mu_i$ ,  $\mu_e$  y  $\mu$  no son observables. Es evidente que no disponemos de todos los valores atribuidos al *universo* por todos los evaluadores, ni de los valores de todos los individuos en la *población* atribuidos por un determinado evaluador. Así, podemos definir la puntuación observada  $X_{ie}$  en términos de estos parámetros.

Por tanto, la demarcación que vamos a llevar a cabo introduce una distinción entre las fases del análisis de la varianza y las que se fundamentan en los conceptos de la TG. El modelo del análisis de la varianza tiene en cuenta las observaciones en las que se supone la existencia de fuentes de varianza. Permite precisar la importancia de cada una de estas fuentes de variación, atribuyéndoles una porción de la varianza total. En este modelo, nada evoca la distinción entre puntuación verdadera y del error. Todas las fuentes de varianza son necesarias en una descripción correcta y completa de la realidad observada.

Como se dijo en el apartado inicial, trataremos de estimar en qué medida se puede generalizar a partir de la media observada en nuestro estudio a la media de todas las observaciones posibles. Para conseguirlo, siguiendo a Blanco-Villaseñor (1993), pasaremos por cuatro fases; las dos primeras estarían relacionadas con la investigación básica, a través del análisis de varianza, mientras que las dos últimas pertenecen a la investigación aplicada y es donde se realiza el estudio de la generalizabilidad.

En una primera fase, puramente descriptiva, trataremos de identificar y organizar los datos en lo que se denomina Plan de Observación. Es el momento de elegir las variables o facetas (en terminología de Cronbach) que nos interesan estudiar y las interrelaciones entre ellas. Además hay que decidir el número de niveles muestreados en cada faceta y a través del análisis de varianza calcular el cuadrado medio de cada fuente de variación del plan utilizado. Así se pueden probar diferentes estructuras de diseño.

En la segunda fase, denominada Plan de Estimación, se elige el modelo de estimación más apropiado, ya sea de efectos aleatorios o mixtos y estará determinado por el modo de muestrear los niveles de cada faceta. Así se puede estimar una faceta a un nivel aleatorio infinito o mixto en el que se combinan los niveles admisibles con los niveles observados, con facetas infinitas, finitas o fijas (es decir, que no se puedan generalizar, ya que los niveles observados agotan los niveles admisibles).

En la tercera fase, se introducen los conceptos de la TG y se realiza el Plan de Medida. Aquí se debe precisar la intención de medida y qué faceta o facetas constituyen el objeto de estudio privi-

legiado. La TG nos permite crear una disimetría entre las facetas, ya que unas van a ser consideradas las fuentes de varianza deseables o verdaderas, debidas a las diferencias entre los objetos de medida, por los que se denomina *Diferenciación*, mientras que las otras serían fuentes de fluctuaciones aleatorias, es decir, varianza de error, debida a la elección de los instrumentos utilizados en la medida, por lo que se denomina *Instrumentación*. Las facetas que se consideren objeto de medida (*diferenciación*), constituyen la población objeto de estudio, mientras que los instrumentos de medida (*instrumentación*) constituyen el universo de generalización. En ambos casos pueden existir facetas aleatorias (infinitas o finitas) o fijas.

En la cuarta fase, denominada Plan de Optimización, tratamos, a través de la información que nos proporcionan los análisis precedentes, de identificar la mejor adecuación posible en los procedimientos de medida. A partir de esta información podemos realizar modificaciones alternativas y/o conjuntas de los planes anteriores: de Observación, de Estimación y/o de Medida. El Plan de Optimización podemos aplicarlo tanto a situaciones de medida orientadas hacia una decisión, como a áreas de investigación orientadas hacia una conclusión.

En concreto, para la investigación que nos ocupa el problema sería: Buscar la calidad de Programas de Evaluación en el contexto específico de los programas de Formación de Padres con el fin de implementar y estandarizar su aplicación masiva. Para ello, planteamos la siguiente hipótesis: El aumento o la disminución del número de niveles de las facetas planificadas en el programa debe permitir minimizar los efectos del azar, de tal forma que nos permita generar la estructura de diseño óptima al Programa de Formación de Padres que se pretende.

#### Método

##### Muestra

El Programa Experiencial para Padres se llevó a cabo en 6 grupos pertenecientes a tres zonas de la isla de Tenerife. El estudio se realizó con grupos de familias multiproblemas, aunque también se pretendía intervenir en una población en la que se combinaran los niveles socioeconómicos y culturales. El grupo de familias multiproblemas estaba compuesto por 25 madres que tenían a sus hijos en un comedor social ubicado en Ofra, zona suburbana de Santa Cruz de Tenerife. Acudieron al programa en dos grupos, uno de mañana (de 17 madres) y otro de tarde (de 6 madres). Estas familias presentaban alto número de hijos, porcentaje elevado de familias monoparentales, niveles educativos y profesionales bajos... Además existían casos de madres e hijos maltratados, consumo de droga por hijos menores, hacinamiento, prostitución...

Los otros grupos estaban compuestos por familias de diferentes niveles socioculturales, mayoritariamente familias nucleares y con menor número de hijos. En Tacoronte, zona semirural al norte de la isla, se formaron dos grupos, uno de mañana (de 6 madres) y uno de tarde (5 madres). Y otros dos en La Matanza de Acentejo, una zona rural, uno de mañana (de 10 madres y 2 padres) y uno de tarde (16 madres y 3 padres).

Se utilizaron dos grupos categóricos, uno para la observación al grupo de padres y otro para observar a los mediadores. Las categorías de los padres tenían una referencia directa a diferentes dimensiones del comportamiento de los mismos y hacían referencia a la utilización de términos nuevos, a los contenidos alternativos

a su complejidad, a la identificación de la acción (Razonamiento y Consecuencias), a las creencias (Globales y Praxis), al nivel de expresión (Impersonal, Personal y Autojustificación) y al clima (Conflicto, Agresivo y Participación). Las categorías del guía mediador hacen una referencia explícita al tema (Centrar, Exponer ideas, Aspectos relevantes y Aportar conclusiones), al clima (Experiencia compartida, Interrogatorio, Conflicto, Agresivo), a la ayuda a las demandas, al estilo participativo (Anima, Propone, Interrumpe, Contradice, Negocia) y a la gestión del tiempo (Participa, No participa, Descontrol, Reduce discusión). La información adicional sobre los constructos teóricos utilizados en la definición de las categorías, así como el Manual del Observador, que describe cada una de las categorías de ambos sistemas se encuentra en los trabajos de Máiquez (1997) y Máiquez, Rodrigo, Capote y Vermaes (2000).

##### Diseño del plan de observación

El Programa consta de 4 Módulos repartidos en un total de 12 sesiones o reuniones. Tanto los mediadores como los observadores, recibieron un entrenamiento previo sobre sus actuaciones; el mediador sobre los principios del programa y su misión en el grupo de padres, y los observadores sobre las categorías a observar en los padres y en el mediador.

Con la evaluación continua, como se explicitó anteriormente, se pretende llevar un control sobre el desarrollo de las sesiones, que permita, al contemplar el proceso seguido, poder introducir modificaciones y tener los datos necesarios para valorar si se han conseguido los objetivos propuestos. En la observación al grupo de padres nos interesaba detectar el proceso seguido de acuerdo con los objetivos del programa. Así se elaboraron 13 categorías relacionadas con la utilización de términos, de contenidos, sobre el nivel de identificación de la acción, creencias manifestadas, nivel de expresión y clima grupal. En la observación al guía-mediador, pretendíamos detectar las pautas seguidas por el profesional durante las sesiones y poder realizar las modificaciones oportunas si éstas se alejaban del modelo para el que habían sido entrenados. En esta ocasión se trataba de 18 categorías agrupadas de acuerdo con cuatro factores: formas de abordar el tema, clima generado en el grupo, ayuda necesaria a las demandas, estilo participativo y gestión del tiempo.

En cada grupo, un observador realizaba el seguimiento al grupo de padres y al mediador en todas las sesiones. Contaba con dos hojas de observación (padres y mediadores) y cada 20 minutos registraba cómo había evolucionado la sesión durante ese período de tiempo, obteniendo, así, datos en tres momentos (1<sup>o</sup>, 2<sup>o</sup> y 3<sup>o</sup>) de la sesión y en cinco niveles: *nada, poco, algo, bastante y mucho*. Habría que señalar que además de contar con un observador fijo para cada grupo de padres, otro observador, que denominamos *itinerante* acudía como mínimo a dos de las sesiones de cada grupo durante el desarrollo del programa. De esta forma se obtuvo una medida de la fiabilidad interobservadores, contemplada en uno de los posibles Planes de Medida, plan que obviamos, dado los altos valores de precisión obtenidos tanto en las sesiones exploratorias previas como en el resto del registro.

Diseñamos así el Plan de Observación que comprende 4 facetas: Grupos con 6 niveles, Reuniones con 12 niveles, Categorías (13 para Padres y 18 para mediadores) y número de Observaciones con 6 niveles. El plan de observación es multifaceta totalmente cruzado G x R x C x O.

Los resultados del Plan de Observación se obtienen a través del programa 8V del paquete estadístico BMDP (Dixon, Brown, Engelman, & Jennrich, 1990). El programa ofrece una información inicial de suma de cuadrados y cuadrados medios de la tabla resumen del análisis de varianza. La tabla 1 refleja el Plan de Observación G x R x C x O tanto para la observación al grupo de padres como a los mediadores (Tabla 1).

#### Diseño del plan de estimación

Los resultados del Plan de Estimación también se obtienen a través del programa 8V del paquete estadístico BMDP (Dixon, Brown, Engelman, & Jennrich, 1990). La tabla 2 describe, para ambas observaciones, los componentes de varianza y su respectiva contribución en porcentajes que nos proporciona información sobre los componentes de varianza aleatorios  $N_g = N_r = N_c = N_o = \infty$  (Tabla 2).

Tanto en la observación al grupo de padres como al mediador, se aprecian porcentajes de variación imperceptibles para las face-

tas reuniones (R) y observaciones (O), así como para la mayor parte de las interacciones de primer y segundo orden en las que están implicadas estas facetas. El efecto contrario, puede observarse en las facetas grupo (G) y sobre todo en la faceta categorías (C), que aportan un alto porcentaje de la variabilidad total del diseño de investigación en el Plan de Estimación aleatorio. Lo mismo ocurre en algunas de las interacciones de primer y segundo orden en que están implicadas estas facetas.

Esta fase proporciona datos muy interesantes. Ya sabemos que las facetas reuniones y observaciones, al no presentar diferencias significativas no deberían constituir, individual o conjuntamente las *facetas de diferenciación*, aunque obtendríamos buenos resultados si estas facetas constituyeran la generalización. La *diferenciación* de las facetas grupos y categorías, al igual que en la observación a los padres, tendrán una buena precisión ya que las diferencias serán muy significativas y tanto los grupos como las categorías tenderán a diferenciarse unos de otros. Intentaremos corroborar esta información en la fase de optimización del programa.

Tabla 1  
Resumen del análisis de varianza del Plan de Observación G x R x C x O en las observaciones a los padres y a los mediadores

Fuente de variación	Padres			Mediadores		
	Suma de cuadrados	GL	Cuadrados medios	Suma de cuadrados	GL	Cuadrados medios
GRUPOS (G)	351.260	5	70.252	287.130	5	57.426
REUNIONES (R)	23.060	11	2.096	38.590	11	3.508
GR	139.210	55	2.531	152.290	55	2.769
CATEGORÍAS (C)	3478.890	12	289.908	9933.280	17	290.193
GC	644.570	60	10.743	795.380	85	9.357
RC	107.160	132	0.812	142.170	187	0.760
GRC	633.840	660	0.960	796.140	935	0.851
OBSERVACIONES (O)	2.570	2	1.285	13.790	2	6.895
GO	3.510	10	0.351	6.260	10	0.626
RO	4.930	22	0.224	6.820	22	0.310
GRO	17.090	110	0.155	26.060	110	0.237
CO	27.260	24	1.136	56.550	34	1.663
GCO	44.100	120	0.367	86.690	170	0.510
RCO	46.120	264	0.175	70.950	374	0.190
Residual	213.740	1320	0.162	331.560	1870	0.177

Tabla 2  
Componentes de varianza aleatorios y su contribución en porcentajes correspondiente al Plan de Observación G x R x C x O en las observaciones a los padres y a los mediadores

Fuente de variación	Padres		Mediadores	
	Componentes aleatorios	%	Componentes aleatorios	%
GRUPOS (G)	0.1238	5.6	0.0711	3.4
REUNIONES (R)	-0.0015	0.0	0.0024	0.1
GR	0.0404	1.8	0.0344	1.6
CATEGORÍAS (C)	1.2896	59.2	1.2953	62.1
GC	0.2660	12.2	0.2270	10.8
RC	-0.0090	0.0	-0.0058	0.0
GRC	0.2661	12.2	0.2247	10.7
OBSERVACIONES (O)	0.0001	0.0	0.0039	0.1
GO	-0.0001	0.0	0.0003	0.0
RO	0.0007	0.0	0.0006	0.0
GRO	-0.0005	0.0	0.0033	0.1
CO	0.0105	0.4	0.0158	0.7
GCO	0.0171	0.7	0.0277	1.3
RCO	0.0021	0.1	0.0021	0.1
Residual	0.1619	7.4	0.1773	8.5

## Resultados

## Plan de Medida

Con la información obtenida en el Plan de Estimación se determinaron los diferentes Planes de Medida, que permitirán posteriormente llevar a cabo las optimizaciones pertinentes. Se han considerado cinco Planes de Medida: 1) RCO/G ¿Es necesario aumentar el número de grupos para generalizar con mayor precisión?, donde las facetas situadas a la izquierda constituyen la *diferenciación*, mientras que las situadas a la derecha conforman la *generalización* o precisión de la estimación. 2) GCO/R ¿Han sido suficientes las reuniones seleccionadas para generalizar con mayor precisión?. 3) CO/GR ¿Han sido suficientes los grupos y las reuniones simultáneamente para generalizar con mayor precisión?. 4) GRO/C ¿Es necesario aumentar el número de categorías, dada la heterogeneidad y amplitud de las mismas, para generalizar con mayor precisión?. 5) GR/CO ¿Cuántas observaciones por categoría serían necesarias para validar el registro?

## Plan de Optimización

Una vez determinado el Plan de Medida, pasamos a la fase de Optimización, en el que, como se observa en las siguientes tablas, vamos a ir modificando el plan de observación original para lograr una optimización de cada una de las facetas en estudio. Para cada Plan de Medida, realizaremos las modificaciones oportunas en las facetas consideradas como instrumentos de medida. Todas las optimizaciones se han llevado a cabo teniendo en cuenta el Plan de Estimación totalmente aleatorio  $N_g = N_r = N_c = N_o = \infty$  y a través del programa Etudgen (Duquesne, 1986).

El Plan de Medida 1 RCO/G, tanto para la observación al grupo de padres como para la observación al mediador (Tabla 3), nos ofrece un coeficiente de generalizabilidad excelente, que observamos, aumenta a medida que añadimos más grupos (por ejemplo en la observación al mediador, con 20 grupos conseguiríamos un co-

eficiente de 0.974). Pero para conseguir este aumento en la precisión de la generalización de la faceta grupos necesitaríamos realizar un número muy grande de registros (12960 de los 3888 iniciales). Esto nos confirma que, también en este caso, los grupos iniciales, son suficientes para realizar una adecuada valoración de los registros, por lo que no haría falta modificaciones en el diseño ni en el plan de estimación.

El plan de medida 2 GCO/R, ofrece, en ambas observaciones (Tabla 4), un coeficiente de generalizabilidad, aún mejor que el del plan 1 (0.978). En este caso, también se ha disminuido el número de reuniones para comprobar si los coeficientes siguen siendo altos, al tiempo que reducimos el número de registros. Se observa que con 3 ó 4 reuniones se reduce considerablemente el número de registros y seguimos manteniendo un buen coeficiente. Por tanto, tampoco es necesario observar a los padres y a los mediadores en todas las reuniones ya que es posible obtener unos buenos resultados observándolos en 3 ó 4 reuniones de las 12 programadas.

El plan de medida 3 CO/GR, ofrece también un coeficiente de generalizabilidad alto (Tabla 5). En este caso también se ha disminuido el número de grupos y reuniones para comprobar si se consiguen unos buenos coeficientes, al tiempo que se reduce el número de registros. Así podemos observar cómo con 4 reuniones en 3 grupos o con 6 reuniones en 6 grupos, se reduce considerablemente el número de registros manteniendo un buen coeficiente de generalizabilidad. Este resultado confirma que no es necesario realizar las observaciones en todos los grupos y las reuniones, como también veíamos en el Plan 2, para obtener la validez del registro. Aunque metodológicamente es posible disminuir los grupos para validar las categorías y los momentos de observación, conceptualmente no sería adecuado si cada grupo es llevado por mediadores diferentes ya que perderíamos información sobre su actuación y no podríamos ir revisando su proceso.

El plan de medida 4 GRO/C, ofrece unos coeficientes de generalizabilidad algo más bajo que los anteriores (Tabla 6), que con-

Tabla 3  
Optimización de los Planes de Medida en la observación al grupo de padres y al de mediadores: Plan 1

PLAN DE OPTIMIZACIÓN 1			RCO/G		PADRES		
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	2	3	4	5	20
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 13$	$N_c = \infty$	13	13	13	13	13
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 2808		936	1404	1872	2340	9360
	Rho. Relativo= 0.912		0.776	0.839	0.874	0.897	0.972
	Rho. Absoluto= 0.899		0.749	0.817	0.856	0.882	0.967
PLAN DE OPTIMIZACIÓN 1			RCO/G		MEDIADORES		
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	2	3	8	10	20
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 18$	$N_c = \infty$	18	18	18	18	18
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 3888		1296	1944	5184	6480	12960
	Rho. Relativo= 0.919		0.792	0.851	0.938	0.950	0.974
	Rho. Absoluto= 0.912		0.775	0.838	0.932	0.945	0.972

seguimos mejorar aumentando el número de categorías. El problema que se presenta es que con un número elevado de categorías, conseguimos un aceptable coeficiente de generalizabilidad pero a costa de un incremento notorio de registros. Por ello, pensamos que con 25 categorías anidadas en subcategorías que reflejen diferentes dimensiones del comportamiento, se podría obtener una buena información sobre un comportamiento específico de los padres y del mediador en las sesiones sin incrementar tanto el número de registros.

El plan de medida 5 GR/CO, nos ofrece unos coeficientes de generalizabilidad también algo bajos (Tabla 7), que conseguimos mejorar aumentando el número de categorías ya que

el incremento de las observaciones no parecen ser sensibles a cambios. Aunque aumentemos el número de registros en la observación, si no se aumentan las categorías los coeficientes no mejoran. Esto lo podemos ver en la observación al mediador, cuando con 25 categorías y 10 registros de observación, obtenemos un coeficiente de 0.848, mientras que con 35 categorías y 2 registros el coeficiente aumenta, 0.857. Así, manteniendo el número de registros o incluso con uno menos, pero aumentando las categorías podemos mejorar la validez del registro.

En resumen, en lo que respecta a la observación al grupo de padres y al mediador, se obtiene información *a priori* que nos in-

Tabla 4  
Optimización de los Planes de Medida en la observación al grupo de padres y al de mediadores: Plan 2

PLAN DE OPTIMIZACIÓN 2		GCO/R		PADRES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	2	3	4	5	6
<b>C</b>	$n_c = 13$	$N_c = \infty$	13	13	13	13	13
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 2808		468	702	936	1170	1404
	Rho. Relativo= 0.978		0.879	0.916	0.935	0.948	0.956
	Rho. Absoluto= 0.978		0.879	0.916	0.935	0.948	0.956
PLAN DE OPTIMIZACIÓN 2		GCO/R		MEDIADORES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	2	3	4	6	8
<b>C</b>	$n_c = 18$	$N_c = \infty$	18	18	18	18	18
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 3888		648	972	1296	1944	2592
	Rho. Relativo= 0.978		0.881	0.918	0.937	0.957	0.967
	Rho. Absoluto= 0.978		0.881	0.917	0.937	0.957	0.967

Tabla 5  
Optimización de los Planes de Medida en la observación al grupo de padres y al de mediadores: Plan 3

PLAN DE OPTIMIZACIÓN 3		CO/GR		PADRES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	2	2	3	4	5
<b>R</b>	$n_r = 12$	$N_r = \infty$	3	4	4	4	3
<b>C</b>	$n_c = 13$	$N_c = \infty$	13	13	13	13	13
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 2808		234	312	468	624	585
	Rho. Relativo= 0.961		0.859	0.869	0.909	0.930	0.938
	Rho. Absoluto= 0.946		0.821	0.832	0.881	0.908	0.920
PLAN DE OPTIMIZACIÓN 3		CO/GR		MEDIADORES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	2	2	3	6	10
<b>R</b>	$n_r = 12$	$N_r = \infty$	3	4	4	6	6
<b>C</b>	$n_c = 18$	$N_c = \infty$	18	18	18	18	18
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 3888		324	432	648	1944	3240
	Rho. Relativo= 0.965		0.870	0.880	0.917	0.960	0.976
	Rho. Absoluto= 0.956		0.847	0.857	0.900	0.951	0.970

dica que se parte de un diseño de investigación adecuado que nos permite generalizar nuestros resultados, y que puede ser mejorado incrementando las categorías de observación. Pero este incremento debe ser moderado, es decir, que nos permita obtener un buen coeficiente sin realizar un gran número de observaciones. Teniendo en cuenta la aplicabilidad del programa pueden conseguirse buenos resultados incrementando moderadamente el número de categorías, que serían suficiente registrar en dos momentos de la reunión o sesión. Además sólo sería necesario observar a los padres entre 4 y 6 reuniones, con lo que se mantiene una buena información reduciendo el coste del programa en personal observador.

### Discusión

Como se vio en el primer apartado de este artículo, aunque diferentes autores coinciden en señalar que una valoración paso a paso permite el cambio de cualquier aspecto en el que no se está consiguiendo el objetivo propuesto (Fernández y Secades, 1999; Harman y Brim, 1980; Secades, 1997), son escasos los programas que realizan un seguimiento del proceso. Como se defiende en el trabajo de Kaplan (1980), la evaluación debe considerarse como una parte del desarrollo del programa de intervención, no un proceso aplicado después de su desarrollo. En lo que ya no existe tanto acuerdo es en los aspectos que deben ser evaluados o en los ins-

*Tabla 6*  
Optimización de los Planes de Medida en la observación al grupo de padres y al de mediadores: Plan 4

PLAN DE OPTIMIZACIÓN 4		GRO/C		PADRES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 13$	$N_c = \infty$	20	25	28	30	35
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 2808		4320	5400	6048	6480	7560
	Rho. Relativo= 0.748		0.820	0.851	0.865	0.872	0.889
	Rho. Absoluto= 0.516		0.621	0.672	0.697	0.711	0.742
PLAN DE OPTIMIZACIÓN 4		GRO/C		MEDIADORES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 18$	$N_c = \infty$	15	20	25	30	35
<b>O</b>	$n_o = 3$	$N_o = \infty$	3	3	3	3	3
	Total observaciones= 3888		3240	4320	5400	6480	7560
	Rho. Relativo= 0.756		0.720	0.775	0.811	0.838	0.857
	Rho. Absoluto= 0.514		0.469	0.541	0.595	0.638	0.673

*Tabla 7*  
Optimización de los Planes de Medida en la observación al grupo de padres y al de mediadores: Plan 5

PLAN DE OPTIMIZACIÓN 5		GR/CO		PADRES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 13$	$N_c = \infty$	13	13	20	30	30
<b>O</b>	$n_o = 3$	$N_o = \infty$	2	10	6	2	1
	Total observaciones= 2808		1872	9360	8640	4320	2160
	Rho. Relativo= 0.782		0.773	0.795	0.853	0.886	0.874
	Rho. Absoluto= 0.531		0.526	0.537	0.639	0.719	0.711
PLAN DE OPTIMIZACIÓN 5		GR/CO		MEDIADORES			
Facetas	Niveles observados	Niveles estimados	1	2	3	4	5
<b>G</b>	$n_g = 6$	$N_g = \infty$	6	6	6	6	6
<b>R</b>	$n_r = 12$	$N_r = \infty$	12	12	12	12	12
<b>C</b>	$n_c = 18$	$N_c = \infty$	18	25	25	35	45
<b>O</b>	$n_o = 3$	$N_o = \infty$	12	3	10	2	3
	Total observaciones= 3888		155	5400	18000	5040	9720
	Rho. Relativo= 0.781		0.803	0.829	0.848	0.857	0.893
	Rho. Absoluto= 0.510		0.522	0.588	0.601	0.654	0.714



trumentos necesarios para dicha evaluación. Kaplan (1980) sugiere que estas informaciones se podrían obtener a través de escritos o diarios de los participantes, observaciones sobre la interacción de padres e hijos en casa, los logros de los participantes evaluados al final de cada sesión o unidad, medición de destrezas y actitudes obtenidas en cada sesión y, como en nuestro caso, a través de los datos que nos proporcionan observadores no participantes sobre el proceso y cómo se lleva a cabo. Pero la realidad es que, en la mayoría de los trabajos que evalúan el proceso, tratan de obtener únicamente datos sobre la satisfacción de los padres con el programa o sobre la participación de éstos en el grupo (Ameijeiras, 1997; Boutin y Durning, 1991; Clarke-Stewart y Fein, 1983; Harman y Brin, 1980). En nuestro programa, estamos satisfechos con la evaluación continua ya que ha supuesto un enriquecimiento para la tarea del mediador y permite seguir la dinámica del grupo de acuerdo con los objetivos propuestos (Máiquez et al., 2000).

Pero también el proceso evaluativo puede realizarse abordando la eficiencia del programa, para lo cual se puede tener en cuenta la

duración del mismo, el tiempo de la sesión, el número de sesiones y el diseño de las mismas (Clarke-Stewart & Fein, 1983; DeShon, Ployhart & Sacco 1998). En nuestro caso, la evaluación de la eficiencia, teniendo en cuenta los resultados que hemos obtenido en el estudio de la generalizabilidad, nos ha servido para comprobar que partimos de un diseño de investigación adecuado en cuanto al número de grupos que han realizado el programa, el número de observaciones y registros realizados y el número de categorías. Pero además, nos ha servido para poder diseñar una intervención aplicada a gran escala, introduciendo modificaciones en el programa actual. En concreto, debemos incrementar el número de categorías de observación, aunque de forma moderada para que no implique la realización de un gran número de observaciones durante las sesiones. Además, manteniendo una buena información, podemos reducir el coste del programa en personal observador ya que sólo sería necesario observar a los padres entre 4 y 6 reuniones de las 12 programadas. Y se simplificaría el trabajo de observación ya que sería suficiente registrar en dos momentos de la reunión o sesión.

### Referencias

- Algina, J. & Keselman, H.J. (1997). Testing repeated measures hypotheses when covariance matrices are heterogeneous - revisiting the robustness of the Welch-James test. *Multivariate Behavioral Research*, 32 (3), 255-274.
- Ameijeiras, S. (1997). *Escuelas de padres y madres*. Comunicación presentada en el VI Congreso Internacional de Educación Familiar. Benalmádena (Málaga) 30 Abril- 4 Mayo.
- Anguera, M.T. (1990). Programas de intervención. ¿Hasta qué punto es factible su evaluación? *Revista de Investigación Educativa*, 8 (16), 77-93.
- Anguera, M.T. (1992). Incidence of cost-benefit analysis and cost-effectiveness analysis in Program Administration. A case study. In J. Mayne, M.L. Bemelmans-Videc, J. Hudson & R. Conner (Ed.), *Advancing Policy and Program Evaluation: Learning from International Experiences* (pp. 181-188) Amsterdam: Elsevier.
- Anguera, M.T. (1995). Diseños. En R. Fernández Ballesteros (Ed.), *Evaluación de Programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 149-172). Madrid: Síntesis.
- Anguera, M.T. (1996). Evaluación de intervenciones en situaciones aplicadas. En D.E. Gómez y X.L. Saburido (Eds.), *Salud y prevención. Nuevas aportaciones desde la Evaluación Psicológica* (pp. 185-213). Santiago de Compostela: Universidad de Santiago de Compostela.
- Arruabarrena, M.I. y De Paül, J. (1995). Los programas de tratamiento de familias con problemas de maltrato y abandono infantil: Descripción y evaluación. *Infancia y Aprendizaje*, 71, 159-178.
- Berk, R.A. (1979). Generalizability of behavioral observations: A clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency*, 83, 460-472.
- Blanco Villaseñor, A. (1992). Aplicaciones de la Teoría de la Generalizabilidad en la selección de diseños evaluativos. *Bordón*, 43 (4), 431-459.
- Blanco Villaseñor, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M.T. Anguera (Ed.), *Metodología observacional en la investigación psicológica* (Vol 2 Fundamentación, pp 151-261). Barcelona: PPU.
- Blanco-Villaseñor, A., Losada, J.L., y Anguera, M.T. (1991). Estimación de la precisión en diseños de evaluación ambiental. *Evaluación Psicológica / Psychological Assessment*, 7 (2), 223-257.
- Boutin, G. y Durning, P. (1997). *Intervenciones socioeducativas en el medio familiar*. Madrid: Narcea.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Skokie, IL: Rand McNally.
- Cardinet, J. (1997). From classical test theory to generalizability theory - the contribution of ANOVA [french]. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée*, 47 (3), 197-204.
- Cardinet, J., & Tourneur, Y. (1985). *Assurer la mesure*. Berne: Peter Lang.
- Clarke-Stewart, K.A. & Fein, G.G. (1983). Early childhood programs. En P.H. Mussen (Ed.) *Handbook of child psychology*. Vol. II. New York: Wiley.
- Coleman, K.J. & Epstein, L.H. (1998). Application of generalizability theory to measurement of activity in males who are not regularly active - a preliminary report. *Research Quarterly for Exercise & Sport*, 69(1), 58-63.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- Crocker, P.R.E., Bailey, D.A., Faulkner, R.A., Kowalski, K.C. & McGrath, R. (1997). Measuring general levels of physical activity - preliminary evidence for the physical activity questionnaire for older children. *Medicine & Science in Sports & Exercise*, 29 (10), 1344-1349.
- Cronbach, L.J., Rajaratnam, N., & Gleser, G.C. (1963). Theory of generalizability: a liberalization of reliability theory. *British Journal of Mathematical and Statistical Psychology*, 16, 137-163.
- Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Cronbach, L.J., Linn R.L., Brennan, R.L. & Haertel E.H. (1997). Generalizability analysis for performance assessments of student achievement or school effectiveness. *Educational & Psychological Measurement*, 57 (3), 373-399.
- DeShon, R.P., Ployhart, R.E. & Sacco, J.M. (1998). The estimation of reliability in longitudinal models. *International Journal of Behavioral Development*, 22 (3), 493-515.
- Dixon, W.J., Brown, M.B., Engelman, L., & Jennrich, R.I. (1990). *BMDP Statistical Software Manual*. Berkeley, CA: University of California Press.
- Duquesne, F. (1986). Développement sur micro-ordinateur d'un programme pour l'étude de la généralisabilité des données. *Scientia Paedagogica Experimentalis*, 23 (1), 29-36.
- Fernández-Ballesteros, R. (1995). *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.
- Fernández Hermida, J.R. y Secades Villa, R (1999). La evaluación de programas de tratamiento para drogodependientes en España. *Psicothema*, 9 (2).
- García Cueto, E. (1996). Aplicación de modelos psicométricos para tests y medidas alternativas en Psicometría. *Psicothema*, 8 (2).
- Gerin, W., Christenfeld, N., Pieper, C., Derafael, D.A., Su, O., Stroessner, S.J., Deich, J. & Pickering, T.G. (1998). The generalizability of cardiovascular responses across settings. *Journal of Psychosomatic Research*, 44 (2), 209-218.

- Gracia, E. (1997). *El apoyo social en la intervención comunitaria*. Barcelona: Paidós.
- Harman, C. & Brim, O.G. Jr. (1980): *Learning to be parents: principles, programs and methods*. Beverly Hill, CA: Sage Publications.
- Irving, P.G., Coleman, D.F. & Cooper, C.L. (1997). Further assessments of a three-component model of occupational commitment - generalizability and differences across occupations. *Journal of Applied Psychology*, 82 (3), 444-452.
- Kaplan, M.S. (1980). Evaluating parent education programs. En M.J. Fine (Ed.) *Handbook on parent education* (pp. 317-334). New York/London: Academic Press.
- Li, M.N.F. & Lautenschlager, G. (1997). Generalizability theory applied to categorical data. *Educational & Psychological Measurement*, 57 (5), 813-822.
- Máiquez, M.L. (1997). *Programa de enriquecimiento experiencial para padres*. Tesis Doctoral no publicada. Universidad de La Laguna. Tenerife.
- Máiquez, M.L., Rodrigo, M.J., Capote, C. y Vermaes, I. (2000). *Aprender en la vida cotidiana: un programa experiencial para padres*. Madrid: Visor.
- Marcoulides, G.A. (1997). Optimizing measurement designs with budget constraints - the variable cost case. *Educational & Psychological Measurement*, 57 (5), 808-812.
- Mcbee, M.M. & Barnes, L.L.B. (1998). The generalizability of a performance assessment measuring achievement in eighth-grade mathematics. *Applied Measurement in Education*, 11 (2), 179-194.
- Mitchell, S.K. (1979). Interobserver Agreement, Reliability, and Generalizability of Data Collected in Observational Studies. *Psychological Bulletin*, 86 (2), 376-390.
- Murawski, M.M. & Miederhoff, P.A. (1998). On the generalizability of statistical expressions of health related quality of life instrument responsiveness - a data synthesis. *Quality of Life Research*, 7 (1), 11-22.
- Nattinger, A.B., McAuliffe, T.L. & Schapira, M.M. (1997). Generalizability of the surveillance, epidemiology, and end results registry population - factors relevant to epidemiologic and health care research. *Journal of Clinical Epidemiology*, 50 (8), 939-945.
- Plewis, I. (1988). Estimating generalizability in systematic observation studies. *British Journal of Mathematical and Statistical Psychology*, 41, 53-62.
- Rindermann, H. (1997). Generalizability of student ratings of university teaching - can the evaluation results of one course be transferred to other courses of the same university teacher or to courses with equal content but different teachers. *Psychologie in Erziehung und Unterricht*, 44 (3), 216-234.
- Rossi, P.H. & Freeman, H.E. (1989). *Evaluation: A Systematic Approach* (4<sup>th</sup> ed). Newbury Park, CA: Sage.
- Secades Villa, R. (1997). Evaluación conductual en prevención de recaídas en la adicción a las drogas: Estado actual y aplicaciones clínicas. *Psicothema*, 9 (2).
- Shavelson, R.J., & Webb, N.M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications.
- Smith, P.L., & Teeter, P.A. (March, 1982). *The Use of Generalizability Theory with Behavioral Observation*. Paper presented at the Annual Meeting of the American Educational Research Association. New York.
- Sun, A.J., Valiga, M.J. & Gao, X.H. (1997). Using generalizability theory to assess the reliability of student ratings of academic advising. *Journal of Experimental Education*. 65 (4), 367-379.
- Tome, M.B. & Isaac, M.T. (1997). Cost-benefit and cost-effectiveness analysis of the rapid onset of selective selective reuptake inhibitors by augmentation. *International Journal of Psychiatry in Medicine*, 27 (4), 377-390.
- Vila, I. (1998). Intervención psicopedagógica en el contexto familiar. En M.J. Rodrigo y J. Palacios (Coords.), *Familia y desarrollo humano*. Madrid: Alianza.

Aceptado el 18 de abril de 2000