

Estimación de la precisión en la evaluación de un programa de intervención familiar

María Luisa Máiquez Chaves*, Ángel Blanco Villaseñor** y Carmen Capote Cabrera*

* Universidad de La Laguna, ** Universidad de Barcelona

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, 1999). 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 of a program for family intervention. The present study is aimed to show the efficiency estimation model based on the «Experiential Program for Parents» (Máiquez, Rodrigo, Capote y Vermaes, 1999). 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. All these meeting were especially useful not only for parents but also for the assistant-mediator to be advised, thereby analysing the whole process. Generalisability theory allow 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.

Cualquier intento de minimizar o eliminar un problema social implica una teoría, un programa y usualmente una gran cantidad de dinero. La eficacia de los programas para reducir dichos problemas son todos ellos problemas de la evaluación. Dado que los programas son importantes para la Comunidad puesto que deben proporcionar beneficio social y además son muy costosos presupuestariamente la evaluación de los mismos se considera prioritaria.

La otra cara de la moneda, la conceptualización de la investigación evaluativa se inicia con la publicación del trabajo de Campbell y Stanley (1966) sobre los diseños de investigación utilizados en la evaluación de programas educativos. Se crea así un nuevo vocabulario para la taxonomización de los diseños de investigación y para la discusión de los principios de la validez de los mismos.

La investigación evaluativa fue inicialmente la quintaesencia de la evaluación de los efectos netos de los programas y el principal problema en el diseño de investigación fue especificar las condiciones apropiadas que permitían estimaciones válidas de estos efectos netos. Las investigaciones empíricas realizadas en el ám-

bito de los diseños experimentales con su máxima expresión de aleatorización, de manipulación y de control no fueron la solución para la investigación evaluativa, dado los sesgos de aleatorización y de control interno debilitado que presentaban las mismas, además del concepto de manipulación que se transformaba en intervención con el fin de evaluar el impacto que producían.

Campbell y Stanley (1966) intentan solucionar este dilema, acuñando el término cuasi-experimental con el fin de cubrir el ámbito de la investigación evaluativa donde la aleatorización y el control presentaban sesgos de consideración. Aunque ellos mismos reconocen la falta de validez generada por este tipo de datos, también expresan las condiciones bajo las cuales se pueden realizar inferencias válidas desde los estudios evaluativos que usen tales tipos de diseños (Cook y Campbell, 1979); y sobre todo porque éticamente es imposible utilizar la aleatorización en la evaluación de la mayor parte de los programas.

Sin embargo, la evaluación de programas adolece de los mismos sesgos planteados por Campbell y Stanley, al tratarse de estudios aplicados que han de implementarse posteriormente con el fin de producir beneficios sociales. Por ello, el análisis de la información a través de un estudio de generalizabilidad ha sido diseñado con el fin de producir estimaciones de componentes de variancia «a priori» que podrán ser utilizados «a posteriori» para preparar más eficientemente investigaciones aplicadas que contribuyan a un beneficio social. El trabajo empírico que presentamos trata de sol-

Correspondencia: Ángel Blanco Villaseñor
Facultad de Psicología
Universidad de Barcelona
08035 Barcelona (Spain)
E-mail: ablanco@psi.ub.es

ventar dichos problemas a través del Programa Experiencial para Padres desarrollado en centros comunitarios de Tenerife (Máiquez, 1997; Máiquez, Rodrigo, Capote y Vermaes, 1999).

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 en una intervención familiar que se han beneficiado del mismo, la distribución de las sesiones de observación, 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. Asimismo, pueden realizarse estudios económicos en términos de la relación coste-beneficio del programa, lo cual es muy útil cuando se busca su implantación a gran escala.

El objetivo de este estudio es presentar el modelo de evaluación de la eficiencia seguido con el Programa Experiencial para Padres en centros comunitarios de Tenerife. 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 et al. (1972) asumen 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 generabilidad, 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.

Por tanto, nuestros objetivos netos serán: 1) 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, y 2) Minimizar los errores de muestreo generados por la utilización de facetas no-aleatorias en los diseños cuasi-experimentales.

La teoría de la generalizabilidad (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 y Anguera, 1991). 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.

Muestra y Estructura de Diseño

El Programa Experiencial para Padres se llevó a cabo en 6 grupos pertenecientes a tres zonas de la isla de Tenerife (Tabla 1): grupos de familias multiproblemas, compuesto por madres que tenían a sus hijos en un comedor social ubicado en Ofra, zona suburbana de Santa Cruz de Tenerife. Los otros grupos estaban compuestos por familias de diferentes niveles socioculturales, en Tacoronte, zona semirural al norte de la isla, y en La Matanza de Acentejo, una zona rural. En los tres casos en grupos de mañana y tarde. El Programa consta de 4 módulos repartidos en un total de 12 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.

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 reuniones 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 reuniones. Contaba con dos hojas de observación (padres y mediadores) y cada 20 minutos registraba cómo había evolucionado la reunión durante ese período de tiempo, obteniendo, así, datos en tres momentos (1º, 2º y 3º) de la sesión. 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 reuniones de cada grupo durante el desarrollo del programa.

En una primera fase, puramente descriptiva, trataremos de identificar y organizar los datos en lo que se denomina *Plan de Observación* (Tabla 1). 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 variancia calcular el cuadrado medio (CM) 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).

Uno de los procesos esenciales en cualquier estudio de generalizabilidad es la *estimación de los componentes de variancia* (Ilustración 1). Para ello, es necesario calcular el Cuadrado Medio Esperado [CME], es decir el valor del cuadrado medio que hubiésemos obtenido en diferentes análisis de muestras repetidas procedentes de la misma población y universo y en la misma es-

tructura de diseño. Los cuadrados medios esperados proporcionan la suma ponderada de los componentes de variancia (Shavelson & Webb, 1991), en el diseño multifaceta $G \times O \times C$. de la Ilustración 1.

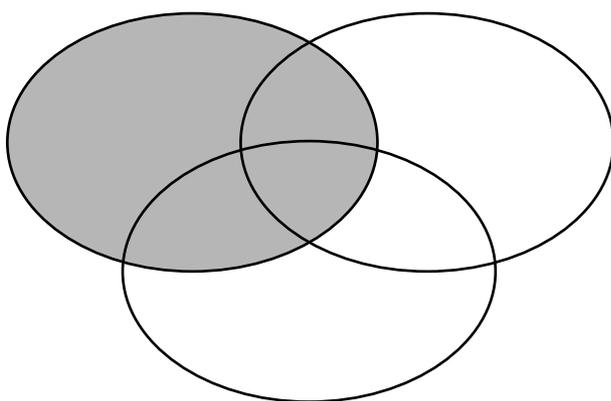
El componente de variancia estimado en un estudio de generalizabilidad refleja la magnitud del error cometido cuando generalizamos el valor obtenido por un individuo a través de un evaluador a la puntuación uni verso (valor medio obtenido a través de los evaluadores en el universo). Luego, los componentes de variancia estimados no reflejan el error al generalizar desde la puntuación atribuida por un evaluador (suma de n_e evaluadores).

La variancia total a través del *universo y población* para este diseño de medida cruzado multifaceta es igual a la suma de los respectivos componentes de variancia de cada uno de los 15 efectos en nuestra estructura de diseño completa Grupos x Reuniones x Categorías x Observadores ($G \times R \times C \times O$) que reflejamos en la Ilustración 2 y que por extensión sería similar al cálculo realizado en la Ilustración 1.

Tabla 1

Plan de Observación de la Estructura de Diseño Multifaceta Totalmente Cruzada correspondiente a la Reunión número 1 (Consideraremos la faceta «Sesión» como una faceta **fija**, dado que han sido seleccionadas para representar una situación estándar y evidentemente no podríamos generalizar a otras situaciones en las que no se cumpliera esta condición)

GRUPOS		REUNION número 1																	
		OFRA Zona suburbana						TACORONTE Zona semi-rural						LA MATANZA Zona rural					
Sesión		Mañana			Tarde			Mañana			Tarde			Mañana			Tarde		
Observación		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
C A T E G O R Í A S	Categorías mediadores																		
	Categorías padres																		



$$CME_g = \hat{\sigma}_{[g]}^2 = \hat{\sigma}_{goc,e}^2 + n_c \hat{\sigma}_{go}^2 + n_o \hat{\sigma}_{gc}^2 + n_o n_c \hat{\sigma}_g^2$$

$$CME_o = \hat{\sigma}_{[o]}^2 = \hat{\sigma}_{goc,e}^2 + n_g \hat{\sigma}_{oc}^2 + n_c \hat{\sigma}_{go}^2 + n_g n_c \hat{\sigma}_o^2$$

$$CME_c = \hat{\sigma}_{[c]}^2 = \hat{\sigma}_{goc,e}^2 + n_g \hat{\sigma}_{oc}^2 + n_c \hat{\sigma}_{gc}^2 + n_g n_c \hat{\sigma}_c^2$$

$$CME_{go} = \hat{\sigma}_{[go]}^2 = \hat{\sigma}_{goc,e}^2 + n_c \hat{\sigma}_{go}^2$$

$$CME_{gc} = \hat{\sigma}_{[gc]}^2 = \hat{\sigma}_{goc,e}^2 + n_o \hat{\sigma}_{gc}^2$$

$$CME_{oc} = \hat{\sigma}_{[oc]}^2 = \hat{\sigma}_{goc,e}^2 + n_g \hat{\sigma}_{oc}^2$$

$$CME_{goc,e} = \hat{\sigma}_{[goc,e]}^2 = \hat{\sigma}_{goc,e}^2$$

Ilustración 1. Plan de Estimación Aleatorio de la Estructura de Diseño Multifaceta Totalmente Cruzada correspondiente a tres de las cuatro facetas que se han tenido en cuenta. En esta situación concreta, y a título ilustrativo, Grupos x Observaciones x Categorías ($G \times O \times C$). La zona de trama refleja que la faceta Grupos ha sido considerada Faceta de Diferenciación, el resto Facetas de Generalización. El componente estimado en TG a través de la suma ponderada refleja la magnitud del error cometido cuando generalizamos el valor obtenido a la puntuación uni verso. Por tanto, el valor de CM que hubiésemos obtenido en diferentes análisis de muestras repetidas procedentes de la misma población y universo y en la misma estructura de diseño

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 o, como proponemos en la Tabla 4, crear

Grupos Categóricos diferentes donde las categorías sean diferentes en cada grupo (anidación), 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 serían suficiente registrar en dos momentos de la reunión.

Referencias

- 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* (pp.151-261). Vol. 2 Fundamentación (2). 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.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Skokie, IL: Rand McNally.
- 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.
- Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally.
- 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.
- 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.
- 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. (1999). *Aprender en la vida cotidiana: un programa experiencial para padres*. Madrid: Visor.
- Shavelson, R.J., & Webb, N.M. (1991). *Generalizability Theory: A Primer*. Newbury Park, CA: Sage Publications.