

# 9

## *Generalización de resultados en evaluación de programas*

*Salvador Chacón Moscoso  
M. Teresa Anguera Argilaga  
Julio Sánchez-Meca*

Uno de los principales fines que se persiguen en la evaluación de un programa de intervención, además de obtener información para valorar y tomar decisiones sobre el programa objeto de la evaluación, es intentar obtener conclusiones acerca de las posibilidades de aplicación de ese programa en otros ámbitos de intervención más o menos similares. Por este motivo la generalización de los resultados evaluativos es uno de los principales aspectos que se deben considerar en evaluación de programas.

Dadas las características del ámbito de aplicación de la evaluación de programas normalmente suelen utilizarse diseños evaluativos de intervención media (cuasi-experimentos) e intervención baja (estudios observacionales/etnográficos). En ambos casos este tipo de evaluaciones suelen estar muy circunscritas a los ámbitos concretos de intervención, de tal forma que usualmente se suele criticar que los resultados evaluativos obtenidos son difíciles de generalizar.

En este contexto el objetivo que se persigue en este capítulo es describir de una forma sistemática y resumida la combinación de distintas dimensiones estructurales de los principales diseños evaluativos al uso que habrían de considerarse desde el inicio del proceso evaluativo para potenciar la generalización de los resultados.

Para conseguir este objetivo en este capítulo vamos a seguir el siguiente proceso: en primer lugar, se va a justificar el análisis desde los elementos implicados en la validez externa y la generalización de resultados de acuerdo con el marco de análisis planteado por Shadish, Cook y Campbell (2002); seguidamente se desglosarán los elementos de diseño que se tienen que considerar en las evaluaciones de intervención alta, media y baja respecto a cada uno de los elementos del acrónimo "UTOSTi" (usuarios, acciones de programa, *outputs*, contexto de implementación y temporalización) tomado como referente para el estudio de la validez externa; y se planteará la convergencia entre las dimensiones de diseño de evaluación y se describirán las consideraciones que hay que tener en cuenta para potenciar la generalización de los resultados evaluativos.

En relación con este análisis se describirá el paralelismo existente entre dichas dimensiones estructurales de diseño y las variables utilizadas para seleccionar, clasificar y analizar los datos en meta-análisis.

En la última parte del capítulo se pondrán en relación los procesos de generalización tanto desde el marco de análisis de la validez, como desde los cinco principios de la teoría de generalización causal de Cook (1990), tomados como criterios referentes desde los que valorar los elementos de diseño considerados.

### 9.1. Validez, generalización y dimensiones estructurales de diseños de evaluación de programas

Al referirnos a la generalización de resultados evaluativos se está planteando la posibilidad de obtener conclusiones generales que puedan ser aplicables a ámbitos de intervención particular. Tal y como se planteó en el capítulo 3 sobre validez en el proceso de la evaluación de programas, la generalización evaluativa debería realizarse sobre los cinco referentes de validez (unidades, tratamientos o acciones del programa, observaciones, contexto y momentos temporales). También en ese capítulo se planteó que se pueden realizar dos tipos de generalizaciones acerca de cada uno de esos referentes de validez (Shadish, Cook y Campbell, 2002). Por una parte nos referimos a cuáles son los constructos poblacionales asociados "UTOSTI" con sus casos concretos de "utosti", utilizados en la evaluación (validez de constructo), y por otra parte se hacía referencia al grado en que los resultados encontrados en nuestra evaluación particular podrían ser extrapolados sobre poblaciones diferenciales "\*UTOSTI" o subconjuntos de "UTOSTI" (validez externa).

A partir de este marco de análisis de la generalización a la hora de diseñar un diseño evaluativo, ya sea experimental, cuasiexperimental u observacional/etnográfico, se deberían considerar desde un primer momento una serie de dimensiones elementales de diseño ligadas a cada uno de los referentes del "utosti". En términos generales se hace referencia a los diseños experimentales cuando cumplen tres condiciones (manipulación de variables independientes –tratamientos, intervenciones–, la medición de los efectos que esta manipulación tiene sobre otras variables consideradas variables dependientes –efectos del programa–, y la asignación aleatoria de los usuarios a las condiciones del programa), los diseños cuasiexperimentales o de intervención media presentan estas mismas características a excepción de la asignación aleatoria de los usuarios a las condiciones del programa. Por otra parte, se considerarán en este capítulo los diseños observacionales como aquellos en los que no se realiza manipulación de variables para observar los efectos, sino aquellos diseños en los que se observa de la manera menos intrusiva posible la ocurrencia espontánea de las variables objeto de la evaluación sin provocar su comportamiento.

La diferenciación entre los diseños de intervención media e intervención baja se perfila adoptando un criterio convencional, aun cuando en la realidad existe una dimensión continua (Tunnell, 1977) de control interno en la cual se ubicaría cada situación específica de evaluación de programas.

El contexto habit  
ubica en una sutil gra  
individuo puede entr  
ejemplo, un program  
medicamento. O pue  
cierto que le permite  
alimenticios de form  
tir periódicamente a  
su comportamiento r  
durante un período t  
ser progresivamente i

La gradación im  
ción a los de media i  
convencionalmente s  
les y observacionales,  
y de baja intervenció

En este marco d  
des de generalizaci  
viamente cuáles van  
estructura inicial ser  
tros resultados. Con  
como dimensiones d  
diciones del prograr  
pecto a los tratamien  
ción que conlleva la  
nivel conforme se de  
de considerar tanto  
mentos utilizados, s  
ceso evaluativo; resp  
condiciones de apli  
modulan o interact  
rían considerar tan  
relevantes en el estu  
planteadas respect

Los UTOSTI co  
como referentes en l  
lación y el estableci

Nos referimos  
ciones, como pue  
incidencia de los a  
a partir de un mo  
intrínsecos o extrín

stante entre dichas dimensiones, clasificar y analizar

procesos de generalización, cinco principios de la teoría y los referentes desde los que

ños

planteando la posibilidad de bits de intervención para el proceso de la evaluación de los cinco referentes de dimensiones, contexto y momento. Se realizarán dos tipos de generalizaciones (Cook y Campbell, 2002): las asociadas a "UTOSTi" (fidelidad de constructo), y por los referidos en nuestra evaluación "UTOSTi" o sub-

de diseñar un diseño evaluativo, se deberían contemplar los elementos de diseño ligados a la validez interna. Se hace referencia a los diseños de variables independientes que esta manipulación de los factores del programa—, y la validez externa, los diseños cuasiexperimentales a excepción de la validez interna. Por otra parte, se consideran los que no se realizan en los que se realiza la manipulación de las variables objeto

de intervención baja se perfecciona una dimensión concreta de la situación específica de

El contexto habitual que caracteriza a los diseños evaluativos de baja intervención se ubica en una sutil gradación que se inicia en la vida cotidiana de un individuo. Pero este individuo puede entrar en un programa que altera muy poco sus conductas habituales; por ejemplo, un programa farmacológico, que únicamente comporta la administración de un medicamento. O puede iniciar un programa dietético de reducción de peso, que si bien es cierto que le permite comer y cenar en su casa, y a su hora habitual, modifica sus hábitos alimenticios de forma desde leve a drástica. Pero puede ocurrir que además tenga que asistir periódicamente a un gimnasio unos días y horas determinados a la semana, con lo cual su comportamiento no es espontáneo, sino habitual, siempre que sea usuario del programa durante un período temporal suficiente, o incluso puede dejar de ser habitual para pasar a ser progresivamente intrusivo.

La gradación imperceptible que se puede establecer desde los diseños de baja intervención a los de media intervención es sumamente relevante, y debe tenerse en cuenta, aunque convencionalmente se dicotomicen los diferentes diseños evaluativos en cuasiexperimentales y observacionales, o, lo que es equivalente, en diseños evaluativos de media intervención y de baja intervención.

En este marco de trabajo, para poder plantearnos con mayores garantías las posibilidades de generalización de los resultados evaluativos encontrados, es necesario reflexionar previamente cuáles van a ser los elementos que van a conformar nuestro diseño. A partir de esta estructura inicial será posible analizar el grado de generalización que se podrán dar a nuestros resultados. Concretamente respecto a los usuarios del programa se habrían de plantear como dimensiones de diseño los criterios de selección y asignación de los usuarios a las condiciones del programa, así como la posible conformación de grupos de comparación; respecto a los tratamientos o acciones del programa se habría de plantear el nivel de intervención que conlleva la implementación del programa, así como los posibles cambios de dicho nivel conforme se desarrolla el programa; respecto a las observaciones/resultados, se habrían de considerar tanto el tipo de datos registrados y la calidad de ese registro como los instrumentos utilizados, sus tipos y los posibles cambios en el uso de instrumentos durante el proceso evaluativo; respecto al contexto de implementación se habrían de considerar tanto las condiciones de aplicabilidad de la intervención, como las posibles variables contextuales que modulan o interactúan con el desarrollo del programa; en la dimensión temporal se deberían considerar tanto el número de registros como los momentos de registro de los datos relevantes en el estudio evaluativo. En el cuadro 9.1 se muestran las dimensiones de diseño planteadas respecto a cada referente del acrónimo "utosti".

Los UTOSTi constituyen indudablemente las dimensiones estructurales básicas que actúan como referentes en los diseños evaluativos. Sin embargo, no hay que olvidar que cabe una modulación y el establecimiento de unos límites en la evaluación de un programa en concreto.

Nos referimos a diversos condicionantes que generan una amplia casuística de situaciones, como puede ser el grado de implicación en el programa, los rechazos puntuales, la incidencia de los allegados al usuario, la necesidad de replantear las acciones del programa a partir de un momento dado, la no periodicidad en la obtención de datos por motivos intrínsecos o extrínsecos al programa, etc.

**Cuadro 9.1. Dimensiones estructurales del diseño en evaluación de programas referida a la validez del sistema de análisis**

1. U	2. Unidades (normalmente, personas)	3. Criterios de selección 4. Criterios de asignación 5. Grupos de comparación
6. T	7. Tratamiento/acciones del programa	8. Nivel de intervención 9. Cambios en el nivel de intervención
10. O	11. Resultados/instrumentos (outcomes)	12. Tipo de dato 13. Calidad del dato 14. Instrumentos 15. Tipos de instrumentos 16. Cambios de instrumentos
17. S	18. Contexto de implementación (setting)	19. Condiciones de viabilidad 20. Variables contextuales moduladoras 21. T <sub>i</sub> 22. Momentos (time) 23. Número de medidas 24. ( $\leq 1$ , $\geq 2$ ) 25. Momentos de medida

Por consiguiente, al planteamiento propio de cada diseño deberán superponerse los necesarios ajustes que permitan contemplar la realidad en la que se ha implementado el programa.

A continuación se pasará a describir las características principales que presentan estas dimensiones estructurales de diseño en las evaluaciones que usualmente se desarrollan en los distintos ámbitos de intervención; es decir, tanto en las evaluaciones de intervención media (diseños cuasiexperimentales) como en las evaluaciones de intervención baja (diseños observacionales/etnográficos).

### 9.2. Hacia la integración de las dimensiones estructurales de los diseños evaluativos de intervención alta, media y baja

En la evaluación de programas de intervención convergen intereses sustantivos y metodológicos. Los intereses sustantivos deben primar en tanto en cuanto deben cumplirse las normas éticas relativas que hay que procurar evitar: cualquier tipo de malestar físico o psíquico

co a cualquier individuo  
sión, al haberse detecta

No obstante, el cur  
dimental adecuado, y  
blecer con las mayores  
bilidades de generalizaci  
se han planteado las pr  
ceso de la evaluación de  
del UTOSTI: usuarios  
mentación y temporal  
soluciones distintas, to  
aumentar tanto el rigo  
de los resultados.

Cada uno de este  
intervención, sea alto  
janzas y diferencias. I  
fil propio, y que en a  
programa, pero que,  
fio; y semejanzas -ló  
dibujan un horizonte  
cada uno de ellos oc  
nes de vecindad se h  
integración.

En el cuadro 9.  
mente, a partir de u  
los niveles de interv  
van disminuyendo.

Se considera qu  
diseños de alta, me  
necesaria. Y lo es po  
percibirse como ur  
demandan unas acc  
dología más adecua  
finalmente, pero no  
plementar diferent  
man el "tránsito" d  
ocasiones en su co

No se nos esca  
en el significado s  
ble, como lo demu  
tos (sanitarios, ser  
gencia de los crite

co a cualquier individuo; y si el programa no actuase se incumpliría la norma ética por omisión, al haberse detectado una necesidad y no ser ésta atendida.

No obstante, el cumplimiento de los intereses sustantivos requiere el desarrollo procedimental adecuado, y, en consecuencia, un planteamiento metodológico que lleve a establecer con las mayores garantías de validez tanto la valoración del programa como las posibilidades de generalización de los resultados obtenidos. En este sentido en el presente capítulo se han planteado las principales dimensiones de diseño que se deben considerar en el proceso de la evaluación de un programa de tal forma que respondan a los grandes componentes del UTOSTI: usuarios, acciones del programa, datos/instrumentos, contexto de implementación y temporalización. Esta descripción nos ha permitido anticipar que, aunque con soluciones distintas, todo diseño evaluativo debería considerar a priori estos referentes para aumentar tanto el rigor de la evaluación concreta como la potenciación de la generalización de los resultados.

Cada uno de estos componentes se ha desarrollado atendiendo al grado o nivel de intervención, sea alto, medio o bajo, pudiéndose observar cómo entre ellos existen semejanzas y diferencias. Diferencias en tanto en cuanto obedecen a planteamientos con perfil propio, y que en alguna medida pueden estar modulados por el ámbito sustantivo del programa, pero que, en esencia, responden a la voluntad de configurar uno u otro diseño; y semejanzas —lógicas por otra parte— porque el conjunto de los diferentes diseños dibujan un horizonte que podría representarse como un imponente continuum en el cual cada uno de ellos ocupa su lugar, que es contiguo de los demás, por lo que estas relaciones de vecindad se hallarán operativizadas en elementos comunes que van a facilitar su integración.

En el cuadro 9.2 se presenta esta idea de continuum mostrándose cómo, efectivamente, a partir de unos mismos referentes de diseño, conforme se va disminuyendo en los niveles de intervención del programa las restricciones metodológicas de distinto tipo van disminuyendo.

Se considera que los principios moduladores desde los que basar la integración de los diseños de alta, media y baja intervención hacen a ésta posible, y todavía más que posible, necesaria. Y lo es por varios motivos: por una parte, porque la evaluación de programas debe percibirse como un proceso unitario e integrado; en segundo lugar, porque los usuarios demandan unas acciones pertenecientes a un programa que se ha de evaluar con la metodología más adecuada, por encima de la diferenciación de modalidades procedimentales; y finalmente, pero no siendo menos relevante, porque en muchos programas convendrá complementar diferentes metodologías, o incluso introducir mutaciones en el diseño que asuman el "tránsito" de una a otra, en respuesta a una realidad cambiante en los usuarios, y en ocasiones en su contexto, a lo largo del proceso de implementación.

No se nos escapa que existen dificultades prácticas en esta integración, especialmente en el significado simbólico de romper una posición metodológica ortodoxa. Pero es posible, como lo demuestra la propia realidad, en la evaluación de programas de diversos ámbitos (sanitarios, servicios sociales, deportivos, educativos, etc.), en donde cabe una convergencia de los criterios que configuran los respectivos diseños.

evaluación  
análisis

acción  
gnación  
paración

cción  
nivel

o

mentos  
trumentos

viabilidad  
xtuales

e)  
didas

medida

erán superponerse los  
e ha implementado el

es que presentan estas  
ente se desarrollan en  
iones de intervención  
intervención baja (dise-

seños evaluativos

sustantivos y metodo-  
ben cumplirse las nor-  
alestar físico o psíqui-

Cuadro 9.2. Dimensiones estructurales en los diferentes diseños de evaluación

			Alta intervención	Media intervención	Baja intervención
U	Unidades	Criterios de selección	Criterio aleatorio (completamente conocido)	Criterio conocido y desconocido	Criterio conocido y desconocido
		Criterios de asignación Grupos de comparación	Criterio conocido Grupos/personas	Conocido/ desconocido Grupos/personas	Conocido/ desconocido Unidades Pluralidad de unidades (personas y conductas)
T	Tratamiento/ acciones del programa	Nivel de intervención	Alto	Medio/bajo	Bajo. La intervención no afecta a la vida diaria (difundir acciones)
		Cambios en el nivel de intervención	Alto/medio	Alto/medio/bajo	Medio/bajo
O	Datos/ resultados (outcomes)	Tipo de dato	Escala	Escala/ordinal	Nominal, ordinal
		Calidad del dato	No se precisa	Relacionado con el descenso de la estandarización de los instrumentos	Alto control de la calidad del dato (acuerdo entre e intraobservadores)
		Instrumentos	Estándar	Principalmente, instrumentos semiestándar	Principalmente, no estandarizados
		Tipo de instrumentos	Tests estandarizados, medidas psicológicas	Instrumentos semiestandarizados	Registro de conducta (incluyendo documentos escritos)
	Cambios de instrumento	Dependiendo del momento (cambios en el diseño)	Dependiendo del momento (cambios en el diseño)	Dependiendo del momento (cambios en el diseño)	
S	Contexto de implementación (setting)	Condiciones de viabilidad	Elevadas restricciones para su uso	Restricciones medias para su uso	Bajas o ninguna restricción para su uso
		Variables contextuales moduladoras	Dependiendo de los programas de intervención	Dependiendo de los programas de intervención	Dependiendo de los programas de intervención

[.../...]

Ti. Momentos (time) Núm med ( $\leq 1$ , Mor. regis

El principio que tico. A la hora de eva polar sus resultados, que diferencian un ti para utilizarla en la e ello fácil de respondi para poder extrapola ámbitos en mayor o pectiva integradora, rigurosas posibles (e sobre la intervención tales y cuasiexperime ser menos restrictiva

Con este plante lizar y desglosar sist que difícilmente po único tipo de diseñ inestables y cambia situaciones híbridas ma situación. Aden den y suelen cambi derar la mutabilidad

9.3. Meta-análisis

En tanto que el ob ciples característi

ños de evaluación

in	Baja intervención
✓	<p>Criterio conocido y desconocido</p> <p>Conocido/ desconocido</p> <p>Unidades</p> <p>Pluralidad de unidades (personas y conductas)</p>
	<p>Bajo. La intervención no afecta a la vida diaria (difundir acciones)</p> <p>Medio/bajo</p>
	<p>Nominal, ordinal</p> <p>Alto control de la calidad del dato (acuerdo entre e intraobservadores)</p> <p>Principalmente, no estandarizados</p>
	<p>Registro de conducta (incluyendo documentos escritos)</p> <p>Dependiendo del momento (cambios en el diseño)</p>
	<p>Bajas o ninguna restricción para su uso</p> <p>Dependiendo de los programas de intervención</p>

[.../...]

Cuadro 9.2. (continuación)

			Alta intervención	Media intervención	Baja intervención
Ti	Momentos (time)	Número de medidas ( $\leq 1, \geq 2$ )	Dependiendo del diseño previo	Dependiendo del programa del diseño y el contexto de intervención	Dependiendo del programa del diseño y el contexto de intervención
		Momentos de registro	Previo, después y durante las medidas del programa	Previo, después y durante las medidas del programa	Recogidas puntuales y seguimiento (entre sesiones de observación o en las sesiones de observación)

El principio que guía este planteamiento de integración es fundamentalmente pragmático. A la hora de evaluar un programa de intervención y considerar la posibilidad de extrapolar sus resultados, ya sean positivos o no, no se plantean cuestiones acerca de los límites que diferencian un tipo de metodología u otra, o cuál es la técnica de análisis más en boga para utilizarla en la evaluación; la pregunta es directa y de sentido común, aunque no por ello fácil de responder: *¿cuál es el procedimiento más adecuado?*, no sólo para evaluar, sino para poder extrapolar o considerar los resultados obtenidos en nuestra intervención a otros ámbitos en mayor o menor medida similares. Responder a esta pregunta lleva a una perspectiva integradora, en la que partiendo de la posibilidad del uso de las metodologías más rigurosas posibles (en tanto que posibilitan un mayor grado de intervención y de control sobre la intervención objeto de evaluación), ligadas a los diseños de evaluación experimentales y cuasiexperimentales, se llega a los planteamientos de unas metodologías que no por ser menos restrictivas en su aplicabilidad conllevan laxitud o falta de rigurosidad.

Con este planteamiento se ha mostrado como los diseños de evaluación se pueden analizar y desglosar sistemáticamente desde unos mismos parámetros; lo que suele suceder es que difícilmente podemos encontrarnos con situaciones puras en las que sólo se utilice un único tipo de diseño de evaluación frente a otros. Más bien, y debido a las características inestables y cambiantes que definen el ámbito de la evaluación de programas, se suelen dar situaciones híbridas en las que se combinan distintas dimensiones de diseño ante una misma situación. Además, estas situaciones, tal y como se ha comentado anteriormente, pueden y suelen cambiar a lo largo del desarrollo del programa, de ahí la necesidad de considerar la mutabilidad de los diseños de evaluación.

9.3. Meta-análisis y generalización de resultados evaluativos

En tanto que el objetivo de este capítulo es hacer una descripción y reflexión sobre las principales características que se deben considerar en el diseño de un programa de evaluación

para potenciar la generalización de sus resultados, se hace necesario referenciar al meta-análisis en tanto que se trata de un procedimiento de análisis que persigue generalizar resultados a partir de la integración de los resultados en un conjunto de estudios previos sobre una determinada temática (Glass, 1976).

El interés específico de este apartado es mostrar cómo el meta-análisis en tanto procedimiento de análisis dirigido a la generalización de resultados, una vez delimitado el problema y objeto de estudio, han de plantearse una serie de criterios para realizar tanto la revisión bibliográfica como la codificación de los estudios seleccionados para iniciar ese proceso de síntesis comentado. En este sentido, se considera relevante resaltar el paralelismo existente entre los criterios usualmente utilizados tanto para la selección como para la codificación de los estudios usados en meta-análisis y las dimensiones estructurales de diseño anteriormente descritas para potenciar la generalización de los resultados.

Desde este planteamiento, en primer lugar se va a hacer una breve descripción de los pasos y la terminología del meta-análisis con objeto de pasar seguidamente a encuadrar y desarrollar en mayor medida la fase del meta-análisis donde se seleccionan y codifican los estudios, de tal forma que se pueda acabar mostrando el paralelismo entre estas fases y las dimensiones estructurales de diseño anteriormente comentadas.

### 9.3.1. Fases de un estudio meta-analítico

En términos generales las revisiones de estudios tanto cuantitativas como cualitativas siguen una serie de pasos que al menos incluyen la formulación del problema, la recolección de datos, la evaluación de esos datos, el análisis e interpretación y la presentación de resultados (Botella y Gambará, 2003; Cooper, 1998; Cooper y Hedges, 1994; Gómez, 1987; Lipsey y Wilson, 2001; Sánchez-Meca y Ato, 1989). Como en cualquier investigación estos pasos son iterativos e interdependientes, en este sentido a continuación se presentan resumidamente las características que definen propiamente el meta-análisis en esta secuencia lógica de tareas comunes a cualquier proceso de investigación.

#### A) Formulación del problema

El objetivo es plantear un problema claro de investigación, que es de sumo interés para las siguientes fases en tanto que será el origen a partir del cual plantear los criterios desde los que tomar decisiones acerca de qué estudios incluir o no en el meta-análisis. En esta primera fase se habrá de realizar una definición conceptual y operativa de los constructos y variables que se pueden considerar, así como la formulación de hipótesis de acuerdo con nuestros intereses de investigación. Ya en esta primera fase se muestra claramente la necesidad de delimitar los constructos de interés, utilizando el esquema de análisis de la generalización correspondiente a los "UTOSTI" (características de usuarios, tratamientos, resultados, contexto y momentos) como criterios sustantivos iniciales desde los que se inicia el meta-análisis.

#### B) Búsqueda de bibliografía

Los estudios incluidos en el meta-análisis. . . anteriormente como n de localización de los (índices de citación, como en fuentes informáticas, informes publicados, informes de último tipo de fuentes en la bibliografía en e y para poder integrar tados encontrados so ría la localización de en bases de datos dis seguros de haber enc

#### C) Codificación de

A la hora de codificar cionó anteriormente unos estudios u otros radoras" que se tien nadas por las decisio

Los meta-análisis fican tanto las caract Stock, 1994). Hasta tivas como a las me tiempo se pueden c Meca, 1997) que a programa y no infl do cómo estas vari en muchos meta-a rísticas del investig cación y la precisio

#### D) Medida de los

Es poco común temática se haga c



### B) *Búsqueda de bibliografía*

Los estudios incluidos en la revisión deben presentar los criterios derivados del objetivo del meta-análisis. Por ello deben considerarse tanto cuestiones sustantivas comentadas anteriormente como metodológicas derivadas de las temáticas de estudio. Los procedimientos de localización de los estudios pueden basarse tanto en fuentes formales de documentación (índices de citación, búsquedas informatizadas, metabuscadores de bases de datos, etc.), como en fuentes informales (colegios invisibles, expertos en el tema, tesis doctorales no publicadas, informes de investigación de proyectos o contratos de investigación, etc.). Este último tipo de fuentes informales son en ocasiones de gran interés, ya que existe un sesgo en la bibliografía en el sentido de publicar trabajos que presentan resultados significativos, y para poder integrar resultados de varios estudios es necesario disponer de todos los resultados encontrados sobre la temática, ya sean significativos o no; por este motivo es necesaria la localización de los denominados documentos informales (no publicados, ni presentes en bases de datos disponibles), a pesar de ello difícilmente se podrá estar completamente seguros de haber encontrado todos los existentes.

### C) *Codificación de los estudios*

A la hora codificar los estudios seleccionados se ha de tener presente, tal y como se mencionó anteriormente, que ya se han considerado una serie de criterios para la inclusión de unos estudios u otros. En este sentido la delimitación de las características o "variables moderadoras" que se tienen que considerar para realizar dicha codificación ya vienen condicionadas por las decisiones anteriores.

Los meta-análisis suelen utilizar procedimientos de codificación comunes para cuantificar tanto las características como los resultados de los estudios (Lipsey y Wilson, 2001; Stock, 1994). Hasta este momento se ha hecho mención tanto a las características sustantivas como a las metodológicas como criterios para codificar los estudios, además al mismo tiempo se pueden considerar las denominadas variables extrínsecas (Lipsey, 1994, Sánchez-Meca, 1997) que aunque no están directamente relacionadas con la implementación del programa y no influyen en los resultados del programa de evaluación, sí que se ha mostrado cómo estas variables están correlacionadas con la magnitud de los efectos encontrados en muchos meta-análisis (las variables extrínsecas usualmente consideradas son: las características del investigador, las circunstancias en que se desarrolló el estudio, la forma de publicación y la precisión en la información publicada del estudio).

### D) *Medida de los resultados de los estudios*

Es poco común que la medida de los resultados de una serie de estudios sobre una misma temática se haga de la misma forma. De ahí que sea necesario convertir los resultados de los

estudios seleccionados a una métrica común antes de poder compararlos. En meta-análisis usualmente se define el resultado de una investigación a partir de las pruebas de significación, dependientes del tamaño de efecto y del tamaño del estudio. Existen distintas formas de medir tamaños de efecto (Fleiss, 1994; Rosenthal, 1994), aunque normalmente suelen utilizarse las siguientes: la diferencia media tipificada ( $d$ ), el coeficiente de correlación de Pearson ( $r$ ), la correlación biserial puntual ( $r_{bp}$ ) y el índice  $Z_r$  de Fisher. A su vez se suelen realizar transformaciones entre ellas dándose transformaciones de "d" a "r" o de "r" a "d" (Sánchez-Meca, 2000).

#### E) Análisis de los resultados

Conceptualmente los datos del meta-análisis sobre tamaños de efecto son analizados como cualquier otro dato en las ciencias sociales y comportamentales, es decir, mediante el uso de estadísticos descriptivos e inferenciales, tanto con técnicas univariadas como multivariadas (Hedges y Olkin, 1985). No obstante existen una serie de características más específicas de los estudios meta-analíticos como son la ponderación de los tamaños de efecto dependiendo del tamaño del estudio, el uso de pruebas de homogeneidad de tamaños de efecto, la naturaleza jerárquica de los datos, la posible existencia de dependencia de los tamaños de efecto de los estudios analizados y la presencia de sesgos en las publicaciones (Shadish, Cook y Campbell, 2002). Dentro de estas peculiaridades se suele dar una secuencia general de tres fases en el análisis e interpretación de los resultados: *a*) análisis descriptivo de las características de los estudios; *b*) estudio de la distribución del tamaño de efecto medio; *c*) búsqueda de variables moderadoras.

#### F) Interpretación y presentación de resultados

En general los resultados del meta-análisis suelen presentar una serie de ventajas, como son la eficiencia al tratar grandes cantidades de información; el rigor científico, al cumplir con las normas de replicabilidad de la investigación; tienen mayor capacidad para detectar efectos pequeños de relevancia social o práctica; presentan mayor potencia estadística al incrementar el tamaño muestral; dan mayor énfasis a la importancia de la magnitud del efecto frente a pruebas de significación; intentan explicar la heterogeneidad de los resultados mediante la búsqueda de variables moderadoras; y suelen ser más sostenibles al basarse en una metodología cuantitativa objetiva sistemática y replicable (Sánchez-Meca, 2000).

A pesar de las ventajas mencionadas, la interpretación y presentación de los resultados presenta principalmente dos problemas: *a*) sesgos ligados a la inclinación teórica del interpretador; *b*) en ciencias sociales, usualmente nos referimos a datos correlacionales, de ahí que las estimaciones de tamaños de efecto de un conjunto de datos en realidad tienen la misma potencia que en los estudios particulares cuando hacemos referencia a meta-análisis en los que se han usado estudios con asignación aleatoria.

#### 9.3.2. Convergencia e y dimensiones

En la primera parte de resultados, se planteó en relación con la clasificación (UTOSTi). Desde los diseños y la necesidad del proceso de evaluación.

En la breve introducción meta-analítica se describen los rasgos característicos del meta-análisis de los estudios correlacionales y se codifican dichos estudios en función de su análisis sobre alguna temática. Los estudios de meta-análisis son los estudios que analizan el efecto de estos estudios en los estudios. Por tanto, los estudios se clasifican por los criterios de selección y clasificación desde la perspectiva de definir qué variables poseen.

En este sentido, se le otorga importancia a la necesidad de considerar las características de los estudios y las características de los estudios entre características seleccionadas de los estudios. De igual manera, se ha planteado la necesidad de definir qué variables se pueden incluir en el análisis, no solo sino también por el hecho de que dicha calidad de los efectos de los estudios (Greenland, 1994; Johnson, Woody y Williams).

Con todo esto se plantea la necesidad de considerar la necesidad de definir la información necesaria para la clasificación de los estudios en función de las características utilizadas para clasificarlos sino tan solo.

9.3.2. *Convergencia entre selección-codificación de estudios en meta-análisis y dimensiones estructurales de diseños de evaluación de programas*

En la primera parte de este capítulo, a partir del esquema de análisis de la generalización de resultados, se plantearon las dimensiones estructurales de diseño o metodológicas a tener en cuenta en relación con cada uno de los elementos sustantivos del nuestro objeto de evaluación (UTOSTi). Desde este análisis se ha justificado la convergencia de las dimensiones de diseños y la necesidad de considerar dichos elementos, no sólo para aumentar la validez del proceso de evaluación, sino para potenciar la generalización de los resultados encontrados.

En la breve introducción sobre los conceptos y fases que se deben considerar en un estudio meta-analítico se ha puesto de manifiesto cómo para poder iniciar los análisis característicos del meta-análisis es necesario delimitar claramente tanto las características sustantivas de los estudios considerados como las condiciones metodológicas para seleccionar y codificar dichos estudios. Es decir, el proceso de revisión desarrollado desde el meta-análisis sobre alguna temática específica se trata de una investigación en la que las unidades de análisis son los estudios empíricos seleccionados, la variable dependiente es la magnitud del efecto de estos estudios y las variables independientes son las características codificadas en los estudios. Por tanto la capacidad de generalización del meta-análisis viene condicionada por los criterios de selección y categorización de los estudios, en tanto será a partir de esta clasificación desde la que sea posible analizar la homogeneidad de los resultados o bien estudiar qué variables podrían explicar la posible heterogeneidad encontrada.

En este sentido autores como Lipsey (1994) o Sánchez-Meca (1997) han dado gran importancia a la necesidad de desarrollar estructuras conceptuales que sirvan para clasificar las características de los estudios que se pueden incluir en el meta-análisis, diferenciando entre características sustantivas, metodológicas y extrínsecas para la selección y codificación de los estudios. De igual forma se han desarrollado trabajos en los que se ha puesto de manifiesto la necesidad de estudiar y valorar o ponderar la calidad de los diseños de los estudios que se pueden incluir en los meta-análisis, no sólo por la necesidad de seleccionar y codificar sino también por su importancia a la hora de desarrollar los análisis posteriores, en tanto que dicha calidad en los estudios se operativiza en una serie de variables moderadoras de los efectos de los programas evaluados (Wortman, 1994; O'Rourke y Detsky, 1989; Greenland, 1994; Jüni, Witschi, Bloch y Egger, 1999; Trichter, 1999; Weisz, Hawley, Pilonis, Woody y William, 2000; Yeaton, Langenbrunner, Smyth y Wortman, 1995).

Con todo esto se puede observar cómo los criterios de selección y codificación son paralelos a los planteados en las dimensiones estructurales de diseño, de ahí que se ponga de manifiesto su consideración para posibles extrapolaciones futuras. De igual forma se pone de manifiesto la necesidad de hacer mención explícita de estas dimensiones, ya que de lo contrario se dificulta mucho el uso de técnicas de síntesis como el meta-análisis al no disponerse de toda la información necesaria y plantearse por tanto serias dudas respecto al sesgo tanto de la selección de los estudios como de los resultados finales encontrados. En este sentido, por tanto, las características utilizadas para seleccionar y codificar los estudios no sólo son importantes para clasificarlos sino también para el posterior desarrollo de los análisis.

Desde el análisis realizado se considera que las dimensiones estructurales de diseño de evaluación planteadas para potenciar la generalización de los resultados siguen una lógica similar o paralela a las características o variables utilizadas en meta-análisis para seleccionar, codificar y ejecutar sus análisis. En este sentido se plantea una complementariedad entre ambas opciones, en tanto que persiguiéndose un mismo fin, la generalización de resultados evaluativos, se está dando un paralelismo entre un planteamiento de diseño y otro de análisis. Todo lo cual no sólo implicaría una mejora de los procesos evaluativos en sí, sino que además posibilitaría el desarrollo de meta-análisis de mayor nivel (O'Rourke y Detsky, 1989).

#### 9.4. Hacia la generalización de los resultados evaluativos

Cualquier responsable en la implementación de programas, o las instituciones que les dan apoyo, se plantea la pregunta de en qué medida cabe una expectativa relativa a la generalización de los resultados obtenidos. Para poder generalizar resultados evaluativos parece claro que debe delimitarse desde dónde se parte y hacia dónde se quiere extrapolar. Hasta este momento se han descrito, desde el marco de análisis de la generalización de los UTOSTI, que se deben considerar una serie de dimensiones de diseño a la hora de abordar un proceso evaluativo que pretenda llegar a resultados generalizables. Se ha planteado la complementariedad de las dimensiones de diseño de intervención alta, media y baja, describiendo incluso el paralelismo existente entre estos elementos de diseño y las variables utilizadas usualmente en estudios meta-analíticos para codificar los tipos de estudios con un mayor o menor grado de calidad en cuanto al diseño que han utilizado.

La solución clásica para lograr representatividad y generalización de resultados ha venido dada desde los planteamientos del muestreo aleatorio (Cochran, 1977). El problema de esta alternativa clásica es que el muestreo aleatorio es difícilmente aplicable a los distintos elementos que se van a considerar en generalización (usuarios, tratamientos o intervenciones, resultados, contexto y dimensión temporal). Incluso en el caso de poder seleccionar aleatoriamente a los usuarios de un programa, éstos no aceptarían formar parte de dicho programa, o incluso en ocasiones no podrían asignarse aleatoriamente a las condiciones de aplicación. Por otra parte, aunque se pudiese aplicar la selección aleatoria, ésta sólo funcionaría parcialmente en tanto que haría referencia a la generalización sobre elementos de la misma población, pero no serviría para realizar generalizaciones sobre otras poblaciones diferenciales.

Desde esta situación el uso de teorías alternativas de generalización son de gran utilidad. En este sentido Cook (1990) planteó formalmente una teoría de validez de construcción sobre generalización causal. Cook planteó cinco principios que pueden facilitar la generalización de inferencias causales obtenidas en diseños de intervención alta y media (experimentos y cuasi-experimentos).

En este capítulo se plantea que los cinco principios de generalización causal planteados por Cook pueden ser aplicados sobre cualquier constructo de interés, no sólo sobre generalizaciones causales. Esta lógica se vincula al análisis de las dimensiones de diseño descritas tanto para los diseños de intervención alta/media como para los de intervención baja. Es decir, a

pesar de las distintas s  
do observar la conver  
rar en el estudio de la  
teado tanto para los d  
de relaciones causales  
lejos de la manipulac

Desde esta lógica  
de generalización, el  
elementos de diseño  
referentes de diseño  
ción es necesario dis  
tes los elementos de

Los cinco princi  
obstante, para hacer  
cripción de estos pri  
palmente en las apo  
integración de los ci

#### A) Principio de sim

Es posible gene  
grama donde los us  
aplicación son simi  
te implementado. L  
zadas en la interven  
características más p

En los diseños c  
laridades y la conse  
determinada en un  
lida en otros usuar  
una red de similitu  
ción de conexiones  
den ser tomados co  
tadas con mayor si

#### B) Principio de h

La generalizac  
ciones encontrada  
con respecto a vari

urales de diseño de eva-  
guyen una lógica similar  
ra seleccionar, codificar  
dad entre ambas opcio-  
resultados evaluativos, se  
le análisis. Todo lo cual  
ue además posibilitaría  
39).

stituciones que les dan  
a relativa a la generali-  
evaluativos parece cla-  
extrapolar. Hasta este  
ción de los UTOSTI,  
de abordar un proce-  
planteado la comple-  
ia y baja, describiendo  
as variables utilizadas  
udios con un mayor o

e resultados ha venido  
) El problema de esta  
a los distintos elemen-  
intervenciones, resul-  
ccionar aleatoriamente  
ho programa, o inclu-  
le aplicación. Por otra  
naría parcialmente en  
na población, pero no  
les.

ón son de gran utili-  
validez de construc-  
eden facilitar la gene-  
ención alta y media

ión causal planteados  
o sólo sobre generali-  
diseño descritas tan-  
ción baja. Es decir, a

pesar de las distintas soluciones de diseño descritas desde uno u otro planteamiento, se ha podido observar la convergencia de ambos desarrollos desde los referentes que se vayan a considerar en el estudio de la generalización. Esta convergencia de dimensiones de diseño se ha planteado tanto para los diseños de alto y medio nivel de intervención, claramente ligados al estudio de relaciones causales, como para los diseños de baja intervención en los que su interés queda lejos de la manipulación de variables para el estudio de relaciones causa-efecto.

Desde esta lógica se considera interesante esbozar esquemáticamente los cinco principios de generalización, en tanto que pueden ser el referente desde el que concretar los distintos elementos de diseño comentados anteriormente. Es decir, además de delimitar cuáles son los referentes de diseño que se deben considerar a la hora de planificar un programa de evaluación es necesario disponer de instrumentos de análisis desde los que valorar si son suficientes los elementos de diseño considerados para potenciar la generalización de resultados.

Los cinco principios de generalización planteados por Cook están interrelacionados. No obstante, para hacer más didáctica su comprensión se van a presentar por separado. Para la descripción de estos principios, además del trabajo original de Cook, vamos a basarnos principalmente en las aportaciones de Shadish (1995a, 1995b), considerando su planteamiento de integración de los cinco principios, así como en la obra de Shadish, Cook y Campbell (2002).

#### A) Principio de similitud proximal

Es posible generalizar con un mayor grado de certeza sobre otras aplicaciones del programa donde los usuarios, tratamientos, resultados, contextos y momentos temporales de aplicación son similares a aquellos utilizados en el programa de intervención originalmente implementado. La estrategia consiste en valorar la conexión entre las operaciones realizadas en la intervención concreta y los objetivos de la generalización, especialmente en las características más prototípicas de los constructos sobre los que se desee generalizar.

En los diseños de baja intervención tiene una traducción directa en la búsqueda de regularidades y la consecuente definición de prototipos. Si se detecta una relación de asociación determinada en un usuario al que se le implementa un determinado programa, y se consolida en otros usuarios, o en otros contextos, o en programas no idénticos, se está tejiendo una red de similitudes proximales, que constituye el primer eslabón hacia una generalización de conexiones entre atributos o variables de interés relacionados. Estos resultados pueden ser tomados como el primer acercamiento a relaciones causales posteriormente detectadas con mayor significatividad.

#### B) Principio de heterogeneidad de las irrelevancias

La generalización, de igual forma, se realiza con un mayor grado de certeza si las relaciones encontradas en nuestro programa de intervención particular se siguen manteniendo con respecto a variaciones irrelevantes en los usuarios, intervenciones, contextos, resultados

y momentos. La estrategia de estudio principal en este caso consiste en detectar las irrelevancias, y donde sea posible incluir una diversidad de éstas en la intervención particular objeto de la evaluación. A diferencia de los diseños evaluativos de intervención alta y media, en los diseños de baja intervención resultan especialmente difíciles las decisiones acerca de dichos límites y de cómo podrías ordenarse en tanto que es justamente a partir de la observación de las conductas que aparecen de forma natural o espontánea de donde parte la delimitación de las variaciones relevantes o no.

*C) Principio de validez discriminante*

La generalización se planteará con un mayor grado de certeza cuando se pueda mostrar que es el constructo objeto de evaluación, y no otro posible elemento, necesario para obtener una relación causa efecto. El procedimiento de análisis consiste en demostrar que el o los efectos se dan sólo con la especificación del constructo dada, y no con otros posibles constructos, sean nuevos o redefiniciones del constructo original objeto de evaluación.

Dado que este principio se apoya en que la generalización es más garantista cuando el objetivo está claramente establecido, y que es necesario hallar unas relaciones causa efecto, este principio se halla ensombrecido en los diseños de baja intervención, ya que el agente causal es difuso, por lo que no se podría hablar de acciones sobre las cuales se ejerza un dominio, sino que se trata de acciones que no alteran la vida habitual de los usuarios.

*D) Principio de interpolación y extrapolación empíricas*

La generalización se plantea con mayor certeza si es posible especificar el rango de variación respecto a las dimensiones que se van a considerar en los usuarios, contextos, tratamientos, resultados y momentos sobre los que las relaciones se muestran fuertes, débiles o no se mantienen en medida alguna. La exploración empírica debe permitir conocer la medida en que se genera una determinada variabilidad a partir de una interpolación producida entre variables contextuales moduladoras, que podría permitir predecir unos resultados semejantes. La extrapolación resulta más compleja, pero resultará sumamente útil cuando se cuente con capacidad predictiva de la relación entre variables relevantes.

*E) Principio de explicación causal*

Se puede generalizar con más certeza desde una evaluación particular de un programa de intervención a otros programas de intervención si se es capaz de especificar completamente las variables relacionadas y los procesos de mediación implicados entre dichas variables. Es decir, es posible aplicar este principio cuando se conocen las variables que se deben poner en relación, el proceso que media entre ellas y sus interacciones. El problema que pre-

sentan los diseños c  
tizar que se cumple  
dictibilidad que de

Tal y como se p  
ro a la denominada  
identifican las mue  
momentos que se h  
lar lo encontrado e  
ciales de usuarios, t  
co principios de ge  
hacen reflexionar s  
rar la definición o  
analizar sistemática  
renciales a los usad

La descripción  
características de  
zación y cómo est  
ponen de manifiest  
luación que fome  
tener presente qu  
también para volv  
zado; es decir, una  
este capítulo, se p  
mismos datos y ap  
para su posible ex

**9.5. Implicaciones**

Normalmente en  
evaluación de pro  
tos, e incluso lleg  
revisando de for  
estaban plantead  
permitió su delin  
su aplicación en

Si nos remon  
poner de maner  
pectiva de conti  
ditación a los cá  
nos atenemos a  
prácticamente h

siste en detectar las irrelevantes a intervención particular de intervención alta y media, es las decisiones acerca de la intervención a partir de la observación de donde parte la delimitación.

cuando se pueda mostrar tanto, necesario para obtener en demostrar que el objeto y no con otros posibles objeto de evaluación. más garantista cuando el agente relaciones causa efecto, intervención, ya que el agente cuales se ejerza un dominio sobre los usuarios.

cificar el rango de variables, contextos, tratamientos fuertes, débiles o permitir conocer la mediana de interpolación producida por unos resultados semejantes útil cuando se cuentan.

particular de un programa es especificar completamente los límites entre dichas variables que se deben controlar. El problema que pre-

sentan los diseños de baja intervención se debe a su casuística, ya que no es posible garantizar que se cumplan estas condiciones debido a las características de escaso control y predictibilidad que definen a la evaluación de los programas de baja intervención.

Tal y como se planteaba anteriormente la generalización implica inferencias ligadas tanto a la denominada validez de constructo como a la externa, es decir, con qué constructos se identifican las muestras de usuarios, tratamientos o intervenciones, resultados, contextos y momentos que se han utilizado en nuestra evaluación, y hasta qué punto se puede extrapolar lo encontrado en la evaluación sobre subconjuntos de poblaciones o constructos diferenciales de usuarios, tratamientos o intervenciones, resultados, contextos y momentos. Los cinco principios de generalización están ligados a estos dos tipos de inferencias, en tanto que hacen reflexionar sobre la forma de delimitar los elementos del "UTOSTI", tanto para depurar la definición o concreción de los constructos a los que nos estamos refiriendo, como para analizar sistemáticamente a qué nos referimos cuando se intenta extrapolar a constructos diferenciales a los usados en el programa de intervención.

La descripción que se ha realizado hasta el momento sobre cuáles son las principales características de diseño que se deben considerar desde los planteamientos de la generalización y cómo estas dimensiones están relacionadas con los principios de generalización, ponen de manifiesto su enorme interés a la hora de planificar y diseñar programas de evaluación que fomenten la generalización de sus resultados. Pero al mismo tiempo se ha de tener presente que esta misma lógica de análisis no sólo puede usarse para planificar sino también para volver a analizar o sacar más partido a un proceso evaluativo una vez finalizado; es decir, una vez finalizada una evaluación, siguiendo la lógica de análisis descrita en este capítulo, se pueden plantear cuestiones de interés para realizar un nuevo análisis de los mismos datos y aportar nuevas conclusiones relevantes, tanto para la evaluación en sí como para su posible extrapolación.

### 9.5. Implicaciones en generalización de resultados evaluativos

Normalmente en la bibliografía al uso se suelen plantear los distintos tipos de diseños de evaluación de programas de una manera diferenciada, descritos como tres eslabones distintos, e incluso llegando a plantearse como encajonamientos estrictos. Tradicionalmente, y revisando de forma exhaustiva la bibliografía existente, los diferentes diseños evaluativos estaban planteados, circunscritos y desarrollados de acuerdo con los moldes clásicos, lo cual permitió su delimitación ajustada y un despliegue metodológico considerable, además de su aplicación en un amplio espectro de intervenciones.

Si nos remontamos a décadas anteriores, hubiese sido impensable o poco probable proponer de manera fundamentada una perspectiva de los diseños evaluativos desde una perspectiva de continuidad que diluye los compartimentos estancos antes existentes. La superación a los cánones establecidos era una norma quizá no escrita, pero infranqueable, si nos atenemos a la bibliografía científica y a los posicionamientos manifestados de manera prácticamente homogénea en eventos científicos. Con el paso de los años, el desarrollo de

estudios evaluativos y la presión creciente que se ejercía por parte de los evaluadores a los "metodólogos", se creó la necesidad de llenar estos intersticios, importantes desde los planteamientos procedimentales clásicos, con el fin de ofrecer una perspectiva más posibilista y dar mayor validez externa a las nuevas propuestas.

En este capítulo se ha dado una visión de continuo en el diseño de programas de evaluación, de tal forma que dependiendo el grado de intervención de la evaluación se han descrito las principales características de los principales tipos de diseños de evaluación. Lógicamente la evaluación desarrollada desde cada una de las tres perspectivas (experimental, cuasiexperimental y observacional) implican características diferenciales. No obstante, se ha puesto de manifiesto cómo no nos estamos refiriendo a compartimentos estancos independientes.

Por el contrario, y dependiendo de las características que presenta el mismo objetivo común, se pretende evaluar y generalizar los resultados obtenidos en un programa de intervención, motivo por el cual habrán de darse las respuestas más acertadas desde la vertiente metodológica que se adecue a dicha demanda. En este sentido se ha planteado un marco global de análisis desde la validez externa, que permite afrontar la problemática de una manera integrada, sin diferenciar entre una perspectiva u otra.

En un primer momento, una respuesta inmediata a la pregunta acerca de qué diseño conviene plantear en una evaluación de programas, puede venir dada por las características del programa, por la naturaleza del contexto, por la forma cómo los usuarios accedieron al programa, por las modificaciones que pudiese sufrir el programa a lo largo del proceso de su implementación, por la naturaleza de los instrumentos que nos han permitido realizar la recogida de datos, por el número de ocasiones en que ha sido posible realizar dicha recogida de datos, etc., o por otros diversos aspectos a los cuales nos hemos referido en este capítulo. Las dimensiones estructurales del diseño recogidas en el UTOSTi constituyen los ejes o referentes por los cuales se pueden regir cualquiera de las pautas básicas que se concretan en un determinado diseño.

Con esta propuesta, probada empíricamente en los últimos años, entendemos que se abre un marco de trabajo prometedor para los evaluadores que han de dar respuestas concretas y eficaces a las situaciones complejas de la realidad de los diferentes ámbitos específicos, y, también para los especialistas en metodología, que siempre deben estar alerta para ofrecer planteamientos y alternativas procedimentales adecuadas en cada caso.

## 9.6. Caso práctico

Generalización de los resultados de un programa de atención a la infancia en comunidades infantiles (adaptado de Chacón, Holgado y Losada, 2006).

Cinco comunidades infantiles acogen a niños en situación de desamparo familiar procedentes de la provincia de Sevilla. Las comunidades se coordinan a partir de un programa global de intervención, a partir del cual se implementan acciones que fomenten el objetivo común de que estos niños y niñas vuelvan a su hogar familiar o a otra familia alternativa.

Grupos de tra  
una de las comuni  
desarrolló un prot  
provinciales y regi  
valoración de obje  
en las cinco comuni  
El proceso imp  
luación; tanto a ni

- a) *Unidades (n*  
ya que los c  
res acogidos  
ción bien de
- b) *Intervencio*  
importante  
vención, en  
a su vez, al  
programa g  
natural se ir  
de interven
- c) *O (resultado*  
siones antes  
en el tiempo  
sionales imp  
sesgo en las  
su evolución  
gidos en sus  
gas a las inici  
es la homog  
ción de los

Todo este proc  
res dentro de cada  
cada programa sin  
comunidades. Tod  
programa global,  
de inclusión/exclu

Desde este an  
necesidad de cons  
evaluación, sino p

El problema c  
vos parece claro c



Grupos de trabajo, utilizando la técnica del grupo nominal, se conformaron en cada una de las comunidades del Servicio de atención a la infancia. Utilizándose esta técnica se desarrolló un protocolo de registro individual con cada menor, y acorde con las directrices provinciales y regionales de atención a la infancia, que consistía en una serie de escalas de valoración de objetivos operativos del programa global que coordinaba las intervenciones en las cinco comunidades infantiles. Este proceso duró tres años.

El proceso implicó una clara delimitación de elementos sustantivos del objeto de evaluación; tanto a nivel poblacional (UTOSTi) como a nivel muestral (utosti).

- a) *Unidades (menores en situación de desamparo)*: nos encontramos en una situación ideal ya que los criterios vienen establecidos legislativamente, y en muchos casos de menores acogidos incluso ratificados por orden judicial. Por tanto, partimos de una población bien delimitada con criterios de selección y asignación, en teoría, homogéneos.
- b) *Intervenciones (treatments)*: merece la pena resaltar que uno de los elementos más importante de este programa es que los datos registrados eran útiles a nivel de intervención, en tanto modulaba los programas de trabajo individual con cada menor, y a su vez, al estar coordinado entre comunidades infantiles facilitaba el rediseño del programa global en su conjunto. Este proceso recursivo favorecía que de manera natural se implementara lo realmente diseñado. Todo ello era acorde con el modelo de intervención provincial y regional andaluz de atención a la infancia.
- c) *O (resultados), S (contexto) y Ti (tiempo)*: cada menor era medido, al menos, en dos ocasiones antes de recibir las intervenciones. Estas medias eran homólogas y se mantienen en el tiempo. Se trata de un programa con una importante participación de los profesionales implicados en la intervención por lo que hay bastante certeza de que no existe sesgo en las medidas. Por otra parte, a todos los menores se les realiza seguimiento de su evolución respecto a los objetivos marcados al inicio de su acogimiento hasta ser acogidos en sus familias iniciales u otra alternativa. Por tanto, se trata de medidas homólogas a las iniciales, previas a la intervención. La gran ventaja que se presenta en este caso es la homogeneidad en los criterios de toma de decisión respecto al desarrollo o derivación de los menores en cada momento particular.

Todo este proceso dio la posibilidad de tomar como grupos de comparación de menores dentro de cada comunidad (al comparar la evolución de cada menor, no sólo dentro de cada programa sino también entre subprogramas); o considerar grupos de comparación entre comunidades. Todo esto es posible, principalmente, por estar implementándose un mismo programa global, por la comparabilidad de las medidas y por utilizar criterios homólogos de inclusión/exclusión de menores en las comunidades.

Desde este análisis se ha justificado la convergencia de las dimensiones de diseños y la necesidad de considerar dichos elementos, no sólo para aumentar la validez del proceso de evaluación, sino para potenciar la generalización de los resultados encontrados.

El problema de la generalización es simple: para poder generalizar resultados evaluativos parece claro que debe delimitarse desde dónde se parte y hacia dónde se quiere extra-

polar. Se ha descrito, desde el marco de análisis de la generalización de los UTOSTi, que se deben considerar una serie de dimensiones de diseño a la hora de abordar un proceso evaluativo que pretenda llegar a resultados generalizables. Se ha planteado la complementariedad de las dimensiones de diseño de intervención alta, media y baja, describiendo incluso el paralelismo existente entre estos elementos de diseño y las variables utilizadas usualmente en estudios meta-analíticos para codificar los tipos de estudios con un mayor o menor grado de calidad en cuanto al diseño que han utilizado.

## 10.1. Introducción

La evaluación de  
económico que p  
cacia obtenida a l  
de las acciones de  
que permiten mo

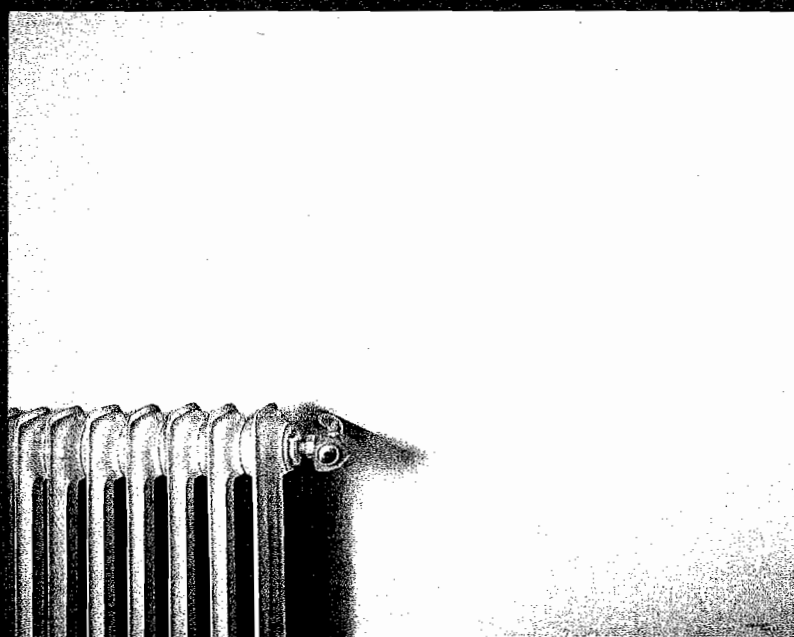
Si un program  
ríamos a él com  
tiempo" (Fontaine  
bubles al prograr  
niencia de imple  
determinadas. Est

El binomio b  
coste de impleme  
pero en torno a es  
ción de programa  
preguntarse de un  
llos programas -s  
o mejorar la calid  
de ellos haya supu

El análisis cui  
ción inicial de los  
de las unidades ob  
costes y los benefi

M. TERESA ANGUERA ARGILAGA  
SALVADOR CHACÓN MOSCOSO  
ÁNGEL BLANCO VILLASEÑOR (coords.)

## EVALUACIÓN DE PROGRAMAS SOCIALES Y SANITARIOS



  
EDITORIAL  
SÍNTESIS

Diseño y óleo de cubierta: Verónica Rubio

Reservados todos los derechos. Está prohibido, bajo las sanciones penales y el resarcimiento civil previstos en las leyes, reproducir, registrar o transmitir esta publicación, íntegra o parcialmente, por cualquier sistema de recuperación y por cualquier medio, sea mecánico, electrónico, magnético, electroóptico, por fotocopia o por cualquier otro, sin la autorización previa por escrito de Editorial Síntesis, S. A.

© M. Teresa Anguera Argilaga, Salvador Chacón Moscoso y Ángel Blanco Villaseñor (coords.)

© EDITORIAL SÍNTESIS, S. A.  
Vallehermoso, 34. 28015 Madrid  
Teléfono 91 593 20 98  
<http://www.sintesis.com>

ISBN: 978-84-9756  
Depósito Legal: M. 34.754-2008

Impreso en España - Printed in Spain