

---

## PONENCIA IV

---

Revista Investigación Educativa - Vol. 8 - n.º 16 - 1990 (P. 77-93)

# PROGRAMAS DE INTERVENCIÓN. ¿HASTA QUÉ PUNTO ES FACTIBLE SU EVALUACIÓN?

*por*

*M. Teresa Anguera Argilaga*

Depto. Metodología de las Ciencias del Comportamiento  
Facultad de Psicología  
Universidad de Barcelona

### RESUMEN

Ante un uso creciente de programas de intervención en todos los ámbitos de la conducta humana, y en concreto en el educativo, resulta urgente y de vital importancia que nos planteemos si no resultará en muchos casos baldío el esfuerzo de elaborar e implementar un programa que presente problemas objetivos intrínsecos que dificulten o anulen la actividad evaluativa, por otra parte necesaria para poder poncernos mediante un juicio de valor sobre la calidad del programa, y, en consecuencia, sobre la relación entre la finalidad perseguida y los efectos resultantes.

Este análisis sobre la factibilidad de la evaluación requiere un punto de partida reflexivo sobre sus propias características definitorias, las cuales se han hallado siempre subordinadas a la propia conceptualización vigente acerca de los programas de intervención, a la vez que sometido a un equilibrio de fuerzas, muchas veces difícil y quizá inestable, entre contenido y procedimiento.

Se trata de sistematizar cuáles son los factores que afectan a la evaluabilidad de un programa y cómo se procede para llevar a cabo ordenadamente dicho análisis, y se concluye con una referencia a sistemas ya existentes al respecto.

### 1. CARACTERÍSTICAS DEFINITORIAS DE UN PROGRAMA DE INTERVENCIÓN: INCIDENCIA DE LAS DISTINTAS CONCEPTUALIZACIONES EN LA FACTIBILIDAD DE LA EVALUACIÓN

Mucho se ha discutido acerca de qué constituye un programa de intervención, cuáles son sus elementos esenciales, o la diversidad «permitida» en su configuración y planteamiento.

Para que sea factible la evaluación de un programa, independientemente de cuál sea el resultado de ésta, y de cómo se haya implementado, deberán cumplirse una serie de requisitos que son los que confieren solidez al propio programa.

La medida en que se hallen presentes unos u otros será en buena parte responsable de que sea factible sin dificultad la evaluación de dicho programa. Ello conlleva, por supuesto, que se parta de un punto de referencia respecto a cómo proceder en y durante la evaluación, así como a la confluencia de todas las estrategias que cumplen un papel más o menos determinante en el transcurso de la evaluación, lo cual debe tenerse presente a modo de cuadro o esquema sistemático abarcador del modelo en su conjunto (Ato, Quiñones, Romero y Rabadán, 1989).

Las diferentes posturas adoptadas oscilan entre la construcción de un corpus de conocimientos que pudiera garantizarle el carácter de disciplina básica, aún en mantillas, y la mera recogida de información que trata de dar cuenta de la ejecución de ciertas actividades realizadas previamente, como un «servicio dedicado a ...» (Cordray, Bloom, & Light, 1987), pasando por el análisis de los efectos netos de los programas o por la influencia de razones políticas al respecto.

Asimismo, muchas son las razones por las que se lleva a cabo una evaluación: Para juzgar el valor de un determinado programa, para estimar su utilidad, para detectar su capacidad de innovación, para incrementar la eficacia de otros tratamientos, para conocer el posible abaratamiento respecto a otras iniciativas de intervención, etc. (Rossi & Freeman, 1986). En cualquier caso, todas ellas se malograrían por inviables si por defecto de construcción del propio programa, o por su carácter incompleto, resultara que estos programas no fueran susceptibles de ser evaluados.

Esta problemática se detecta especialmente cuando contemplamos la evaluación de programas en su devenir histórico. La trayectoria seguida por los estudios sobre evaluación de programas refleja cómo surgió al interés por el tema y las orientaciones hacia las cuales se ha desarrollado, influidas generalmente por las circunstancias sociales y políticas del lugar y momento, dada la diversidad de acciones emprendidas para mejorar no sólo la educación, sino también la salud y las prestaciones sociales en diversos campos.

Un antecedente remoto se halla ya en el siglo XVIII, cuando Thomas Hobbs y sus contemporáneos propusieron sistemas numéricos para evaluar las condiciones sociales de vida de aquel entonces, e indentificando las causas de mortalidad, morbilidad y desorganización social. En la Era Moderna se sitúan en este ámbito los esfuerzos realizados a inicios de este siglo para reducir la mortalidad y morbilidad después de epidemias. Dado su carácter de «evaluación extensiva» (aplicada a grandes masas de población), la mayor debilidad de estos programas residía precisamente en la laxitud de sus presupuestos y su carácter apenas restrictivo respecto a la caracterización de los beneficiarios del programa, lo que equivale a desconsiderar de antemano el posible nexo causal entre la adecuada implementación del programa y la producción de unos resultados.

Fue en torno a 1930 cuando se iniciaron trabajos dirigidos a la evaluación de programas desde una óptica en que predominaban las preocupaciones metodológi-

cas, y con temáticas referidas principalmente a la salud pública (tiempo comprendido entre las dos Guerras Mundiales), creación de empleo, sistemas de propaganda, hábitos alimentarios, y estado de ánimo de los soldados. El período que siguió inmediatamente a la II Guerra Mundial marca el comienzo de programas «a gran escala» orientados al desarrollo urbano, educación tecnológica, prevención sanitaria, entrenamiento ocupacional, planificación familiar y desarrollo comunitario.

Al final de la década de los cincuenta se incrementó considerablemente el interés por este tipo de programas, incorporándose como temáticas nucleares las de prevención de la delincuencia, tratamientos psicoterapéuticos y psicofarmacológicos, rehabilitación de sujetos, encarcelados, además de las propias según zonas (desarrollo agrícola y comunitario en África, nutrición y salud de Latinoamérica, planificación familiar en Asia, etc.). Esta época marca un hito en el interés por incorporar nuevas metodologías y en la utilización de recursos informáticos para el análisis de grandes volúmenes de datos. Este hecho dio lugar a una toma de conciencia acerca de la rigurosidad con que debía plantearse el fundamento de las acciones que constituían el propio programa, y, de aquí, que empezara a tenerse en cuenta la «factibilidad» de la evaluación.

Durante la década de los sesenta, y en Estados de elevado grado de desarrollo económico, se iniciaron una serie de reformas sociales con el objetivo de mejorar la calidad de vida de sus habitantes, que se extienden a los años que medían hasta la actualidad. Se han propuesto el triple objetivo de elaborar programas en base a necesidades concretas, proceder al control del gasto público, y mejorar la calidad de las acciones emprendidas, por lo que se ha ido constituyendo una nueva disciplina, *fundamentalmente aplicada*, que podría titularse «*evaluación de programas de intervención*» (o simplemente «*evaluación de programas*»), a pesar de que ha recibido diferentes denominaciones en función del matiz que se deseaba enfatizar, y que consiste genéricamente en una «aplicación de procedimientos de investigación social al enjuiciamiento o análisis sobre el mérito, valor o eficacia de concretas acciones humanas implantadas en la realidad social» (Fernández Ballesteros, 1987, p. 2). El interés por estudiar la evaluabilidad de los programas es creciente desde este planteamiento, llegándose a considerar como una rama o apartado de especialización dentro del marco de evaluación de programas.

De aquí surge una imperiosa necesidad de estructuración que le confiere un cierto (carácter básico, y por tanto organizado, a esta nueva disciplina (Anguera, 1989; Anguera y blanco, 1988-91), ya que, si partimos, según Cook, Leviton & Shadish (1985) de que un programa es un conjunto heterogéneo de acciones, a veces con una estructura difusa, cuyo objetivo es el de resolver algún problema que incumbe a los ciudadanos, es lógico que se manifiesten muchos fracasos en la aplicación y evaluación de estos programas precisamente por la falta de una estructura sólida y coherente. Y obviamente, si esta disciplina gira en torno a la evaluación de programas, y se le asigna este carácter básico, ello conlleva que, de forma absolutamente imprescindible, se establezcan previamente las pautas, requisitos, filosofía, ... que sustenten su capacidad de ser evaluados.

Precisamente en este punto tal vez convenga recordar a Guba & Lincoln (1987), que en su propuesta de las cuatro generaciones de estilos o estrategias de evaluación, describen la actual (cuarta) como flexible y de carácter colaborativo entre las distintas personas que participan en el proceso en lugar de carices controladores que había tenido anteriormente. Esta filosofía de base permite superar una cierta tensión institucional, y con la ayuda indiscutible de avances de carácter técnico se favorece sin duda a esta nueva configuración como disciplina que se está estructurando.

La evaluación de programas de intervención se puede considerar en la actualidad, en definitiva, como un proceso que intenta determinar, tan estructurada y objetivamente como sea posible, la relevancia, efectividad e impacto de las acciones o tratamientos propuestos a la luz de los objetivos perseguidos (Sohm, 1978), no sin antes implicar sucesivas tomas de posición ante diversos tipos de dilemas (Schalock & Thornton, 1988). Precisamente la forma cómo se han presentado concatenadas las fases de que consta nos muestra la fineza de su desarrollo, cada vez mayor en la actualidad, en que se da rango de fase a cuestiones que antes se hallaban anidadas en otras, como la definición del problema (York, 1982; Mayer, 1985) y la implementación (Williams, 1975; Pancer & Westhues, 1989).

Desde esta perspectiva, se nos muestra cómo necesidad apremiante el conocimiento de la medida en que es *factible* llevar a cabo la evaluación de un programa. Mucho antes que pretender una evaluación exitosa respecto al programa que se está sometiendo a prueba, está la exigencia de disponer de todas aquellas informaciones derivadas del propio programa (Berk & Rossi, 1990), así como de las normas dispuestas respecto a su correspondiente implementación, con el fin de *analizar el propio programa en sí mismo, y por tanto en su estructura (no en sus efectos) una vez que es aplicado a una determinada colectividad.*

Este estudio de «factibilidad» de la evaluación requiere antes tomar en consideración una serie de cuestiones relativas a aspectos diversos, ya que, lógicamente, cuanto mayor sea el rigor con que se someta a prueba la evaluabilidad de un programa, mayor deberá ser la adecuación del análisis a las características específicas y propias del programa considerado.

## 2. BALANCE CONTENIDO-PROCEDIMIENTO

Dentro de la operación previa de análisis del programa en sí mismo, y con la mirada puesta en la medición de su calidad (Moos, 1988), destacamos en primer lugar una revisión de su estructura respecto a un criterio probablemente constatable pero altamente útil a nuestro juicio, y que se refiere al *mayor o menor equilibrio entre contenido y procedimiento*, o, expresado de otra forma, *la evaluación entendida como proceso o como fuente de resultados.*

La característica definitoria de una *evaluación de resultados* se reduce estrictamente a estimar la magnitud de los efectos de un tratamiento sobre el cual se había establecido previamente una hipótesis a modo de conjetura sobre dichos previsibles efectos. La finalidad es claramente demostrativa (Kidder & Judd, 1986), lo que

conlleva ventajas y peligros, como son, por una parte, un claro objetivo funcional rayando en el utilitarismo, que permite conferir un significado muy concreto a la evaluación de programas, pero también, por el lado opuesto, son perfectamente conocidos todos los efectos perniciosos (efecto de la demanda, efecto de experimentador, etc.) que conlleva la intención de contrastar una hipótesis, incluso en colectivos como puede ser en el marco escolar o social (Riecken & Boruch, 1974) que con frecuencia contagian subliminarmente al responsable (evaluador, investigador,... o cualquiera que sea el papel que desempeña en el organigrama de una institución), preocupado esencialmente por la obtención de unos determinados resultados o efectos que en gran parte de los casos han venido precedidos por un esfuerzo importante respecto a la implementación de un determinado tratamiento. Únicamente cuando sea posible desglosar con nitidez el papel del responsable y de los que actúan bajo su supervisión con unos cometidos parciales y de carácter técnico (Barber, 1976) se consigue mitigar la distorsión en el contraste de hipótesis.

Por otra parte, la *evaluación entendida como proceso* tienen unos objetivos muy distintos. En lugar de tratar de demostrar el efecto de un tratamiento, pretende entender por qué se ha producido un determinado resultado. La finalidad no es ya demostrativa, sino explicativa, y en este enclave resulta más cómodo situar la factibilidad de la evaluación. También esta opción alternativa presenta inconvenientes y ventajas. Entre los primeros, y especialmente en el ámbito aplicado, se consideraría probablemente desenfocado el llevar a cabo una intervención (con su correspondiente costo en esfuerzo, tiempo, recursos humanos y materiales,...) con un talante tal que no importara cuál es el resultado. Pero las importantes ventajas que conlleva la evaluación entendida como proceso y que fueron perfectamente expuestas por Judd & Kenny (1981a) son: Al explicar por qué un determinado tratamiento o intervención produce un determinado efecto se incrementa la validez de constructo (Cool & Campbell, 1979; Judd & Kenny, 1981b), la cual se refiere a las relaciones entre las variables y operaciones llevadas a cabo en el estudio empírico y los constructos teóricos que supone que representan.

En este sentido se puede hablar de:

- a) validez de constructo del tratamiento,
- b) validez de constructo de la muestra y del marco de trabajo, y
- c) validez de constructo del resultado o efecto o variable dependiente. Los tres se incrementan a través del proceso de evaluación.

Desde este punto de vista el proceso funciona como una especie de implementación de la evaluación, posibilitando *respectivamente* tres importantes ventajas o beneficios:

- a) El conocer qué componentes cruciales de la(s) variable(s) de tratamiento deban incluirse para contar con suficientes garantías en la generalización de los resultados de la evaluación.

- b) Igualmente importante sería saber de qué tipos de personas se podrían esperar determinados efectos, o en qué situaciones específicas.
- c) Dada la naturaleza compleja e incluso a veces ambigua de muchas de las respuestas que habitualmente interesan, los sujetos aprenden a pretender como resultado respuestas deseables.

Estas dos grandes perspectivas de la evaluación (entendida como resultado y como proceso), en lugar de mostrar su oposición, deben tratar de *complementarse* (Judd, 1987) en aras a un incremento de la interpretabilidad de los resultados, de los componentes de la intervención y de los factores de ámbitos diversos que inciden en cualquier sentido en los efectos alcanzados, lo que podría equivaler a afirmar que gracias a la evaluación entendida como proceso se logra perfeccionar la evaluación entendida como resultado, y, en consecuencia, que se avanza en *evaluabilidad* o factibilidad de una evaluación rigurosa.

Existen además otros argumentos que apoyan la necesidad de complementariedad como forma de superación de debilidades endémicas. Así, por ejemplo, las evaluaciones de resultado se focalizan típicamente en una o varias variables dependientes, que son las presumiblemente afectadas por el tratamiento, pero esto conlleva el riesgo de no poder detectar efectos distintos de los deseados, mientras que si nos situamos desde las evaluaciones de proceso se hallará implícita la inclusión de medidas correspondientes a un amplio elenco de variables que pueden o no hallarse afectadas por el tratamiento, además de tener acceso, a lo largo del seguimiento del proceso, al «comportamiento» de cada una de las variables y en consecuencia qué otras, además de las previsibles según la hipótesis, pueden acusar efectos debidos al tratamiento.

### 3. FACTORES QUE AFECTAN LA FACTIBILIDAD DE LA EVALUACIÓN

El conjunto de factores cuya incidencia debe tenerse en cuenta respecto a la factibilidad de la evaluación de programas de intervención pueden presentarse sistematizados en varios grupos atendiendo a su naturaleza:

#### 3.1. Teórico-conceptuales

Retomando la idea del punto anterior, además de defender la complementariedad entre la evaluación de resultados y de proceso, igualmente habría que partir de la idea de complementariedad entre la consideración de la evaluación como investigación básica y como campo diversificado de trabajo en la práctica diaria.

Si partimos de la premisa del carácter científico de la evaluación (Crane, 1988), éste le confiere unas propiedades que casi podríamos denominar «paradigmáticas» de parsimonia, posibilidad de generalización, coherencia, individualización, especi-

ficación de límites y posibilidad de acumulación de información que tendrán importantes implicaciones respecto a la factibilidad de la evaluación.

Esta cuestión se halla unida a otra que ha sido ampliamente debatida por los teóricos de la evaluación (Bickman, 1987) y que radica en el posicionamiento respecto a la dicotomía probablemente mal llamada «hecho-valor», con raíces profundas en el propio concepto de evaluación, pero también con importantes implicaciones a la hora de precisar los elementos que inciden en la factibilidad de un programa de intervención. Lo que se halla en confrontación es una posición de corte radicalmente positivista o «cientificista» frente a otra de carácter únicamente valorativo.

Autores como Gibbs (1983) realizaron hace ya algún tiempo severas críticas en este sentido. El error (Michalos, 1983) proviene de la confusión de lo que es una distinción ontológica con otra epistemológica. Mientras que la primera posición implica una distinción de significado y uso, la segunda es de confirmabilidad. Las pruebas que puede obtener el educador, o el observador humano en general, implican que se haya partido de hechos contrastables, pero no de juicios de valor. Así, el que haya disminuido el fracaso escolar entre los alumnos de un centro en el que se ha realizado un cambio en los libros de texto utilizados es un «aserto» que deberá ser convenientemente contrastado científicamente a través del uso de una adecuada metodología, pero una mera valoración gratuita de unos y otros libros de texto, por definición subjetiva, es incapaz de probar, y, en consecuencia, de evaluar.

De aquí se derivan puntos de reflexión importantes (Alkin, 1990; Crane, 1988). En primer lugar, la determinación del modelo de probabilidad (clásico, bayesiano,...) implica una larga serie de decisiones/juicios y sólo en unos pocos de ellos existe acuerdo generalizado en el ámbito que aquí nos interesa, como es el educativo y/o de carácter social (no así si se tratara de trabajos experimentales de laboratorio). En segundo lugar, y siguiendo la argumentación de Gibbs (1983), los evaluadores no deben basarse en sus valoraciones personales, sino que deben actuar desde una perspectiva social. Luego los juicios que, siguiendo con el mismo ejemplo, se hubiesen formulado acerca de los distintos libros de texto, utilizando análisis como los de costo beneficio o costo-eficacia, se podrían estandarizar y presentar de forma que no supusiesen problemas epistémicos de orden distinto de los que implica la evaluación de una probabilidad.

De aquí que la medida y la valoración deban actuar complementariamente, siendo cada una de ellas esencial para la otra y habiendo alcanzado igual estatus en el desarrollo de la evaluación como disciplina.

### 3.2. Metodológicos

No es en absoluto nueva la idea de que los evaluadores definen su competencia en función de su formación metodológica (Rossi & Freeman, 1986). Desde una posición analógica, igualmente afirmamos que también los programas de interven-

ción serán tanto más factibles de ser evaluados cuanto mejor se adecúen a los diversos requisitos metodológicos, y éstos, que son inamovibles en lo esencial, a su vez, se pueden estructurar de una forma u otra según características secundarias que deseen enfatizarse (Anguera, Blanco, Gil, Ibern, Losada, Quera, Roca y Redondo, 1989-90).

Si nos circunscribimos al ámbito nacional, en España es justa la crítica de una *endémica debilidad metodológica* de la gran mayoría de los programas que son puestos en práctica por parte de instituciones tanto públicas como privadas<sup>1</sup>. En los países en los que existe una mayor tradición en evaluación de programas (Estados Unidos y Canadá) se aprecian, cada vez de forma más generalizada, *importantes avances consistentes en el uso de recursos metodológicos sofisticados* que permiten un rigor mucho más elevado. A modo de ilustración, podemos señalar en este sentido la aplicación del análisis de supervivencia (Wheeler & Hissong, 1988) y tablas de vida (Benedict, Glasser & Lee, 1989), análisis probabilísticos de evaluación del riesgo (Nichols & Wildavsky, 1988), series de tiempo en cuidadosos estudios longitudinales (Boruch & Pearson, 1988), toma de decisiones (Dawson & Cebul, 1990), análisis de la causalidad en estudios retrospectivos (Holland & Rubin, 1988) y con secuencia temporal de variables (Curry, Yarnold, Bryant, Martin & Hughes, 1988), análisis de discontinuidad en la regresión (Stanley & Robinson, 1990), búsqueda de parámetros en modelos logísticos (Kaplan & Abramson, 1989), análisis logit anidado (Bertrand, Stover & Porter, 1989), estimación de parámetros en diseños de carácter sincrónico-diacrónico (Maxim, 1989), clasificación de modelos mediante análisis de superficie para la búsqueda de dimensiones (Williams, 1989), y estudios de impacto (Ridge, Stern & Watts, 1990), aparte de cuidadosos análisis psicométricos en la construcción de escalas (Alberty & Mihalik, 1989) o de optimizar la complementariedad entre análisis de muy distinto orden (Hennessy & Saltz, 1989). Incluso en evaluaciones planteadas desde un cuño absoluta y/o radicalmente cualitativo, se va apreciando una mejora notable a nivel metodológico, como en cuestiones de conexión de variables en un plano conceptual (Wright, 1989), o de metaevaluación (Cleave-Hogg & Byrne, 1988), o simplemente tratando de incrementar seriamente el rigor en evaluaciones de programas planteadas cualitativamente (Silverman, Ricci & Gunter).

La evaluación es aún una ciencia joven, y ello ha acarreado con frecuencia graves problemas de índole tanto estructural como funcional. Un programa de intervención no puede adoptar el papel de «caja negra», sino que, bien al contrario, debe eliminarse cualquier tipo de opacidad que impida su cabal evaluación. Y es esta opacidad la que deriva en gran medida de un *mal planteamiento del programa* y/o de *irregularidades en su implementación*, así como, en algunos casos, de la *falta de una monitorización adecuada*.

---

1 Hecho constatado muy recientemente a partir de las respuestas recibidas a una solicitud de programas (elaborados y aún no implementados, en curso, y en fase de valoración) dirigida a instituciones de ámbito local, autónomo y nacional en el ámbito de Servicios Sociales (Anguera y Blanco, 1988-91).

### 3.2.1. Defectos de planteamiento del programa

Son de diversa índole, aunque la incorrecta fijación de objetivos es, con mucho, la principal causa que no hace factible la evaluación de un programa. Según Rutman (1980) en un análisis de factibilidad habría que identificar diversos tipos de problemas de los que aquí entresacamos aquéllos que se refieren al planteamiento del programa y añadimos otros que consideramos necesario tener en cuenta:

- a) Programas mal definidos que requieren en consecuencia una posterior elaboración para ser implementados.
- b) Vaguedad en los objetivos, lo cual no permite una suficiente base de sustentación para establecer las directrices del programa.
- c) Fines no realistas o ilusorios, los cuales impedirán un adecuado establecimiento de acciones, ya que su seguimiento sería baldío.
- d) Objetivos contrapuestos, que generarían una contradicción interna en la estructura del programa.
- e) Falta de claridad en el enunciado de las tareas requeridas.
- f) Inexistencia de organigrama relativo a la secuenciación o concurrencia de las distintas acciones de que se compone el programa.
- g) Falta de calendario o previsión temporal.
- h) Imprecisión respecto al estudio empírico, lo cual se diversifica a su vez en uso y validación de instrumentos, naturaleza de los datos y forma de obtención, unicidad de criterios en el almacenamiento de material de archivo, indicadores a utilizar, perspectivas de análisis de datos, etc.

El esfuerzo en resolver estos problemas redundará en un aumento de la factibilidad evaluativa del programa, lo que no quiere decir, por supuesto, que con ello mejore la valoración desde todos los tipos de validez, pues si bien la interna se eleva, pueden en cambio permanecer inamovibles otros tipos, como la externa (sin los resultados no son generalizables) o la social (si la elaboración del programa no estuvo precedida de un análisis previo de necesidades que hiciera a éste pertinente).

### 3.2.2. Irregularidades en su implementación

La implementación (o «proceso», o «administración») de la evaluación no es un concepto nuevo (Freeman, 1977; Finnegan, Murray & MacCarthy, 1988), sino que, por el contrario, cada vez recibe mayor atención (Reichardt & Cook, 1979; Leithwood & Montgomery, 1980; Judd & Kenny, 1981a; Flay, 1986). No obstante, la revaloración de la implementación en la literatura al respecto se ha limitado a señalar su importancia y su papel de componente esencial en la evaluación de los programas de componente esencial en la evaluación de los programas de intervención, pero apenas nadie menciona cómo debería llevarse a cabo (McGraw, McKinlay,

McClements, Lasater, Assaf & Carleton, 1989), por lo que no es de extrañar que no se hayan desarrollado recursos para resolver los programas que pudieran presentarse.

Además, mientras que los defectos estructurales en el planteamiento de un programa son siempre subsanables (y en caso contrario deberá profundizarse en el estudio de la carencia a raíz de la cual se desea intervenir), no ocurre lo mismo a la hora de aplicarlo, es decir, de implementarlo.

Los problemas que pueden surgir en este caso son de dos tipos:

- a) Falta de formación y/o de monitorización en la(s) persona(s) encargada(s) de llevarlo a cabo, lo cual puede ocasionar errores de naturaleza heterogénea. Dada su previsibilidad, se considera grave un fallo ocurrido en las circunstancias indicadas en este apartado, el cual, además, puede tener importantes consecuencias a nivel de costos.
- b) Errores asistemáticos de naturaleza aleatoria, y, en consecuencia, no previsibles: Rechazo de sujetos más allá de los criterios establecidos, incorrecta cumplimentación de protocolos, etc.

El que se mencionen estos dos tipos de problemas no significa que su mera ausencia garantice una elevada «calidad» de la implementación, sobre cuya definición operacional apenas existe consenso. En un trabajo reciente de Pentz, Trebow, Hansen, MacKinnon, Dwyer, Johnson, Flay, Daniels & Cormack (1990) se sugieren al respecto tres elementos que la conforman: 1) *Adherencia*, medida generalmente en función de que un programa se implemente en grupos experimentales y no en grupos control; 2) *exposición*, o parte del programa que ha sido dispensado a los beneficiarios-objetivo; y 3) *reinvención*, o margen en que la implementación se desvía del programa estándar que está siendo sometido a prueba (Fullan & Pomfret, 1977; Pentz & Trebow, press).

De entre estos tres elementos, la adherencia se utiliza principalmente para identificar efectos experimentales y variables contaminantes más que para probar los efectos sobre el comportamiento (Basch, Sliepcevich, Gol, Duncan & Kolbe, 1985). La exposición está siendo estudiada desde el análisis de las inconsistencias en función de la duración del tiempo de implementación, número de sesiones, y longitud de los períodos inter-sesiones (Connell, Turner & Mason, 1985; Tricker & Davis, 1988). Y finalmente la reinvención resulta interesante en la evaluación de las relaciones entre la experiencia de los profesionales que intervienen en la implementación y el hecho de que únicamente se utilicen unas pocas sesiones (Connell & Turner, 1985).

### 3.2.3. Falta de monitorización

No es todavía frecuente institucionalmente el sometimiento del equipo de profesionales que trabajan en la implementación de un programa a un asesoramiento

encaminado a mejorar la implementación, y con ello, en consecuencia, a evitar la desazón que produce la recogida de datos con importante volumen de «missings», o la falta de previsión para evitar la mortalidad experimental, o la duda sobre cómo analizar previamente la validez de un determinado instrumento, o las dificultades para efectuar un análisis de costo-beneficio.

Por el contrario, lo que sí ha llegado a ser habitual en algunos ámbitos es la auditoría, prácticamente retrospectiva por naturaleza, que actúa como de juez imparcial e independiente y que tiene por misión comprobar si la implementación del programa se realiza de acuerdo con el plan previsto (características del programa), a la vez que velar por unas «ratios» adecuadas de costo-beneficio y costo-eficacia. Para llevar a cabo su misión se sirve tanto de estándares ya preparados como de controles específicos (Malan, 1988).

#### 4. NIVELES DE FACTIBILIDAD DE LA EVALUACIÓN

Por los apartados anteriores puede deducirse sin dificultad que es ciertamente compleja la cuestión relativa a la factibilidad de la evaluación, así como múltiples y convergentes las formas de incrementarla. Para Wholey (1983, p. 36), «explora la realidad del programa con el fin de clarificar la plausibilidad de los objetivos del programa y la factibilidad de la medida de los resultados».

Si la tarea de optimización la estructuramos en niveles o etapas, la primera se centraría en la correcta delimitación de fines y objetivos. Con ello, yo podría emprenderse un posterior análisis de documentos, y el lema de Rutman (1980, p. 88) en este sentido es perfecto: «¿Qué aspectos del programa deberían considerarse para su inclusión en una posterior evaluación del programa?». Algunas de las ideas de Rutman (1980) las consideramos para la elaboración de los puntos de referencia en cada una de las fases que abarcaría el proceso de evaluación, y en el cual, consiguientemente, se detectaría la medida en que dicha evaluación es más o menos factible:

A) La primera tarea del evaluador es estudiar la *estructura del programa*, analizando allí las conexiones entre los elementos que lo componen «formalmente» y de forma genérica, y preparando ya con ellos un primer documento que será útil para la evaluación: Institución que lo subvenciona o sponsora, informes anuales, presupuestos previstos para especialistas externos, material necesario para la confección de instrumentos,... La finalidad es un conocimiento de la naturaleza del programa. Para que sea *factible la evaluación* en este primer nivel, es necesario que se disponga de la información pertinente que acabamos de mencionar.

B) Procede el *listado e identificación de los componentes del programa y sus efectos*. En un plano utópico/optimista podría afirmarse que entre ambos existe relación de causalidad, pero la realidad se suele mostrar de forma muy distinta.

Aquí queremos señalar la incidencia de dos importantes aspectos: Prioridades políticas y contexto.

La *dependencia/independencia de políticas* seguidas por las diferentes instituciones que actúan a modo de soporte de determinados programas ha sido un tema polémico, como lo es en general la incorporación de la política a la evaluación (Weiss, 1987; Fernández Ballesteros, 1987).

Mucho se ha discutido sobre la neutralidad del evaluador y su independencia del político, que es quien, previo un análisis de necesidades, deberá tomar decisiones sobre cuestiones tan importantes como el orden de prioridades y el volumen de los recursos que se asignan en cada caso (Rodríguez Castedo, 1987).

Pero además se ha propuesto la existencia de una especie de evaluabilidad de un programa desde una perspectiva estrictamente política, es decir, de una *factibilidad política del programa* (Palumbo, 1987). Así, por ejemplo, se podrá o priorizar entre un programa de alfabetización para niños gitanos u otro de inclusión de maestros de apoyo en las aulas de integración si se diera el caso de que ambos interesan mucho, y el que resulta más adecuado en primer lugar de acuerdo con el momento, las características del barrio, etc., sería el de mayor factibilidad política.

El *contexto* puede llegar a cumplir un importante papel respecto a la producción de efectos, dado el complejo entramado existente de factores interrelacionados (Moos, 1988), por lo que será necesaria la detección del máximo número de conexiones. La situación más clara probablemente sea la del clima social, que siempre incidirá (en uno u otro sentido) sobre los efectos de un determinado programa.

C) Comprobación de la inclusión de *informaciones necesarias para llevar a cabo la implementación* del programa:

- Cómo se implementa.
- Quiénes son los beneficiarios (los que reciben los efectos).
- Quiénes son los usuarios (personas encargadas de la implementación del programa).
- Recursos de diversos órdenes con los que se cuenta.

D) El análisis de la factibilidad requiere una completa especificación de los *requisitos a seguir de índole metodológica*, tanto en lo que se refiere a la construcción de instrumentos, a la obtención de datos y a su tratamiento posterior. Aquí incluimos la indicación del diseño estándar.

E) Existencia de una temporalización detallada (PERT, sistema Ghant) en la que se constaten la secuenciación y concurrencia de las acciones del programa. Precisamente en base a este criterio «de agenda» se han propuesto sistemas de evaluación (Muscatello, 1988).

F) Existen severas limitaciones respecto a la *factibilidad de evaluación de un programa* desde un punto de vista económico: Problemas de elaboración de presu-

puesto, de tasación, de aprovechamiento de bancos de recursos (Drummond, 1980), que alcanza especial gravedad en programas de gran envergadura y de titularidad estatal (Gramlich, 1981). De hecho, de la evaluación de un programa de intervención se espera que se derive una información máximamente objetiva sobre la eficacia y eficiencia del programa mediante las «ratios» correspondientes, así como de los esfuerzos por optimizarlas (Anguera, 1990), y esto conlleva un análisis económico de dicho programa.

G) Adecuada comunicabilidad de la información (clara, concisa, cumplimiento de aspectos formales).

No es de extrañar, en consecuencia, que en los debates que cada vez se plantean con mayor frecuencia en el ámbito evaluativo (Alkin, 1990), y desde los diferentes tipos evaluativos que puedan establecerse (Chen, 1990), se dirijan a los usuarios interesantes sugerencias (Hendricks & Handley, 1990) que recogen ya el conjunto de precauciones a tener en cuenta con el fin de facilitar la actividad evaluadora (Schmidt, Scanlon & Bell, 1989), que no necesariamente será posterior, ya que ello se hallará en función de que el modelo evaluativo seguido sea lineal o no lineal (Veney & Kaluzny, 1984).

## 5. SISTEMAS DE EVALUACIÓN DE LA EVALUABILIDAD

Desde la perspectiva de Wholey (1987), la valoración de la evaluabilidad se materializa en un instrumento diagnóstico y prescriptivo que puede utilizarse para determinar la medida en que perviven los problemas que hemos mencionado.

Como esquema básico, se han propuesto criterios de evaluabilidad (Schipman, 1989), que se presentan en un formato estándar que incide sobre necesidad del programa, implementación y efectos alcanzados, aunque su excesiva generalidad dificulta una actuación objetiva y concreta.

De forma mucho más específica, Fernández Ballesteros (1988) ha elaborado (se halla aún en fase provisional) un instrumento, el *LCREP (Listado de Cuestiones Relevantes en Evaluación de Programas)*, que tiene por objetivo el poder juzgar sobre la dificultad de la evaluación que va a llevarse a cabo, y con la aportación interesante de que cada una de las categorías de análisis se articula en una serie de variables, pudiéndose llegar a una cuantificación de la evaluabilidad en función del cumplimiento de una serie de criterios de calidad establecidos racionalmente.

En el *LCREP* se plantean dos áreas de análisis: En primer lugar, la evaluabilidad del programa en sí misma (centrada en una revisión de las necesidades, objetivos y metas, programa, preevaluación, e implantación), en donde cada uno de los aspectos sometidos a revisión cuenta con cinco puntos referentes a sus principales características, y en base a los cuales se puede asignar una puntuación valorativa. Y una segunda parte se centra en el análisis de las *barreras de la evaluación*, que a su vez se desglosan en el diseño, operaciones observadas, contexto, aceptabilidad de la

evaluación, barreras administrativas y políticas e implicación del evaluador; igual que en la primera parte, sobre cada uno de estos apartados se formulan cinco cuestiones.

En el momento en que se consiga disponer de un instrumento en el que los aspectos sometidos a revisión constituyan una categorización (y por tanto un sistema exhaustivo y mutuamente excluyente) de todas aquellas informaciones necesarias para emprender una correcta evaluación, podremos afirmar que se habrá superado un importante problema endémico. Ahora parece que nos hallemos ya en el borde de su consecución.

## 6. REFERENCIAS

- ALBERTY, S. & MIHALIK, B. J. (1989): The use of importance-performance analysis as an evaluative technique in adult education. *Evaluation Review*, 13 (1), 33-44.
- ALKIN, M. C. (1990): *Debates on evaluation*. Beverly Hills: Sage.
- ANGUERA, M. T. (1989): Innovaciones en la metodología de la evaluación de programas. *Anales de Psicología*, 5 (1-2), 13-42.
- ANGUERA, M. T. (1990, December): *Incidence of cost-benefit and cost-effectiveness analysis in Program Administration*. International Workshop on Policy and Program Evaluation. The Hague, The Netherlands.
- ANGUERA, M. T. y BLANCO, A. (1988-91): *Evaluación de programas en Servicios Sociales: Alternativas metodológicas*. Investigación subvencionada por la Comisión Interministerial de Ciencia y Tecnología (CICYT).
- ANGUERA, M. T., BLANCO, A., GIL, A., IBERN, I., LOSADA, J. L., QUERA, V., REDONDO, S. y ROCA, M. (1989-90): *Instrumentos y técnicas de evaluación de programas de rehabilitación en el ámbito penitenciario*. Investigación encargada por la Dirección General de Servicios Penitenciarios y de Rehabilitación (Consejería de Justicia de la Generalidad de Cataluña).
- ATO, M., QUIÑONES, E., ROMERO, A. y RABADÁN, R. (1989): Evaluación de programas: Aspectos básicos. *Anales de Psicología*, 5 (1-2), 1-12.
- BARBER, T. X. (1976): *Pitfalls in human research. Ten pivotal points*. New York: Pergamon Press.
- BENEDICT, M. B., GLASSER, J. H. & LEE, E. S. (1989): Assessing hospital nursing staff retention and turnover. A life table approach. *Evaluation & the Health Professions*, 12 (1), 73-96.
- BASCH, C. E., SLIEPCEVINCH, E. M., GOLD, R. S., DUNCAN, D. F. & KOLBE, L. J. (1985): Avoiding Type III errors in health education program evaluations: A case study. *Health Education Quarterly*, 12, 315-331.
- BERK, R. & ROSSI, P. H. (1990): *Thinking about program evaluation*. Beverly Hills: Sage.
- BERTRAND, J. T., STOVER, J. & PORTER, R. (1989): Methodologies for evaluating the impact of contraceptive social marketing programs. *Evaluation Review*, 13 (4), 323-354.
- BICKMAN, L. (Ed.) (1987): *Using program theory in evaluation*. San Francisco: Jossey-Bass.
- BORUCH, R. F. & PEARSON, R. W. (1988): Assessing the quality of longitudinal surveys. *Evaluation Review*, 12 (1), 3-58.
- CHEN, H. T. (1990): *Theory-driven evaluations*. Beverly Hills: Sage.
- CLEAVE-HOGG, D. & BYRNE, P. N. (1988): Evaluation of an innovation in a traditional medical school: A metaevaluation. *Evaluation & the Health Professions*, 11 (2), 249-271.
- CONNELL, D. B. & TURNER, R. R. (1985): The impact of instructional experience and the effects of cumulative instruction. *Journal of School Health*, 55, 324-331.

- CONNELL, D. B. & TURNER, R. R. & MASON, E. F. (1985): Summary of findings of the school health education evaluation: Health promotion effectiveness, implementation, and costs. *Journal of School Health*, 55, 316-321.
- COOK, T. D. & CAMPBELL, D. T. (1979): *Quasi-Experimentation: Design and analysis issues for field settings*. Skokie, Ill.: Rand McNally.
- COOK, T. D., LEVITON, L. C. & SHADISH, W. R. (1985): Program evaluation. In G. Lindzey & E. Aronson (Eds.) *Handbook of Social Psychology* (pp. 699-777), 3rd. edition. New York: Random House.
- CORDRAY, D. S., BLOOM, H. S. & LIGHT, R. J. (Eds.) (1987): *Evaluation practice in review*. San Francisco: Jossey-Bass.
- CRANE, J. A. (1988): Evaluation as scientific research. *Evaluation Review*, 12 (5), 467-482.
- CURRY, R. H., YARNOLD, P. R., BRYANT, F. B., MARTIN, G. J. & HUGHES, R. L. (1988): A path analysis of medical school and residency performance. Implications for housestaff selection. *Evaluation & the Health Professions*, 11 (1), 113-129.
- DAWSON, N. V. & CEBUL, R. D. (1990): Advances in quantitative techniques for making medical decisions. The last decade. *Evaluation & the Health Professions*, 13 (1), 37-62.
- DRUMMOND, M. F. (1980): *Principles of economic appraisal in health care*. Oxford: Oxford University Press (Traducido en Madrid: Instituto de Estudios Laborales y de la Seguridad Social, 1983).
- FERNÁNDEZ BALLESTEROS, R. (1987): *Ciencia, ideología y política en evaluación de programas*. Conferencia pronunciada en el curso «Evaluación de programas y servicios para el bienestar social: Un enfoque pluridisciplinario». Santander: Universidad Internacional Menéndez Pelayo (31 agosto - 4 septiembre).
- FERNÁNDEZ BALLESTEROS, R. (1988): *Listado de cuestiones relevantes en evaluación de programas* (LCREP). Manuscrito no publicado. Madrid.
- FINNEGAN, J. R., MURRAY, D. M., KURTH, C. & McCARTHY, P. (1988): Measuring and tracking education program implementation: The Minnesota Heart Health Program. *Health Education Quarterly*, 16 (1), 77-90.
- FLAY, B. R. (1986): Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.
- FREEMAN, H. E. (1977): The present status of evaluation research. *Evaluation Studies Review Annual*, 2, 17-51.
- FULLAN, M. & POMFRET, A. (1977): Research on curriculum and instruction implementation. *Review of Educational Research*, 47, 335-397.
- GIBBS, L. (1983): Evaluation researcher: Scientist or advocate? *Journal of Social Service Research*, 7, 81-92.
- GRAMLICH, E. M. (1981): *Benefit-cost analysis of government programs*. Englewood Cliffs, N.J.: Prentice-Hall.
- GUBA, E. G. & LINCOLN, Y. S. (1987): The countenances of fourth-generation evaluation: Description, judgment, and negotiation. In D. J. Palumbo (Ed.). *The politics of program evaluation* (pp. 202-234). Beverly Hills: Sage.
- HENDRICKS, M. & HANDLEY, E.A. (1990): Improving the recommendations from evaluation studies. *Evaluation and Program Planning*, 13, 109-117.
- HENNESSY, M. & SALTZ, R. F. (1989): Adjusting for multimethod bias through selection modeling. *Evaluation Review*, 13 (4), 380-399.
- HOLLAND, P.W. & RUBIN, D. B. (1988): Causal inference in retrospective studies. *Evaluation Review*, 12 (3), 203-231.
- JUDD, C. M. (1987): Combining process and outcome evaluation. In M.M. Mark & R.L. Shotland (Eds.).

- Multiple methods in program evaluation* (pp. 23-41). San Francisco: Jossey-Bass.
- JUDD, C. M. & KENNY, D. A. (1981a): Process analysis: Estimating mediation in treatment evaluation. *Evaluation Review*, 5, 602-619.
- JUDD, C. M. & KENNY, D. A. (1981b): *Estimating the effects of social interventions*. New York: Cambridge University Press.
- KAPLAN, E. H. & ABRAMSON, P. R. (1989): So what if the program ain't perfect? A mathematic model of AIDS Education. *Evaluation Review*, 13 (2), 107-122.
- KIDDER, L. & JUDD, C. M. (1986): *Research methods in social relations*. New York: Holt, Rinehart & Winston.
- LEITHWOOD, J. & MONTGOMERY, D. J. (1980): Evaluating program implementation. *Evaluation Review*, 4 (2), 193-214.
- MALAN, R. M. (1988): Local government evaluation in a legislative environment. In C. G. Wye & H. P. Hatry (Eds.) *Timely, low-cost evaluation in the public sector* (pp. 21-35). San Francisco: Jossey-Bass.
- MAXIM, P. S. (1989): The impact of design effects on standard errors in roadside traffic surveys. *Evaluation Review*, 13 (2), 157-173.
- MAYER, R. R. (1985): *Policy and program planning*. Englewood Cliffs, N. J.: Prentice-Hall.
- McGRAW, S. A., McKINLAY, S. M., McCLEMENTS, L., LASATER, T. M., ASSAF, A. & CARLETON, R. A. (1989): Methods in program evaluation. The process evaluation system of the Pawtucket Heart Health Program. *Evaluation Review*, 13 (5), 459-483.
- MICHALOS, A. (1983): Technology assessment, facts and values. In P. Durbin & F. Rapp (Eds.) *Philosophy and technology* (pp. 59-81). Dordrecht: Reidel.
- MOOS, R. H. (1988): Assessing the program environment: Implications for program evaluation and desing. In K. J. Conrad & C. Roberts-Gray (Eds.) *Evaluating Program Environments* (pp. 7-23). San Francisco: Jossey-Bass.
- MUSCATELLO, D. B. (1988): Developing and agenda that works: The right choice at the right time. In J.A. McLaughlin, L. J. Weber, R. W. Covert & R. B. Ingle (Eds.) *Evaluation utilization* (pp. 21-31). San Francisco: Jossey-Bass.
- NICHOLS, E. & WILDAVSKY, A. (1988): Regulating by the numbers: Probabilistic risk assessment and nuclear power. *Evaluation Review*, 12 (5), 528-546.
- PALUMBO, D. J. (1987): Politics and evaluation. In D.J. Palumbo (Ed.) *The politics of program evaluation* (pp. 12-46). Beverly Hills: Sage.
- PANCER, S. M. & WESTHUES, A. (1989): A developmental stage approach to program planning and evaluation. *Evaluation Review*, 13 (1), 56-77.
- PENTZ, M. A. & TREBOW, E. A. (in press): Implementation issues in drug abuse prevention reserach In C. Leukfeld & J. Ludford (Eds.) *NIDA Research Monograph*.
- PENTZ, M. A. & TREBOW, E. A., HANSEN, W. B., MacKINNON, D. P., DWYER, J. H., JOHNSON, C. A., FLAY, B.R., DANIELS, S. & CORMACK, C. (1990): Effects of program implementation on adolescent drug use behavior. The Midwestern Prevention Project (MPP). *Evaluation Review*, 14 (3), 264-289.
- REICHARDT, C. S. & COOK, T. D. (1979): Beyond qualitative versus quantitative methods. In T. D. Cook & C. S. Reichardt (Eds.) *Qualitative and quantitative methods in evaluation research* (pp. 7-33). Beverly Hills: Sage (Traducido en Madrid: Morata, 1986).
- RIDGE, R. S., STERN, G. A. & WATTS, R. K. (1990): Econometric model evaluation. Implications for program evaluation. *Evaluation Review*, 14 (3), 308-314.
- RIECKEN, H. W. & BORUCH, R. F. (1974): *Social experimentation. A method for planning and evaluating social intervention*. New York: Academic Press.

- RODRÍGUEZ CASTEDO, A. (1987): *La evaluación en Servicios Sociales*. Conferencia pronunciada en el curso «Evaluación de programas y servicios para el bienestar social: Un enfoque pluridisciplinario». Santander: Universidad Internacional Menéndez Pelayo, 31 agosto - 4 septiembre.
- ROSSI, P. H. & FREEMAN, H. E. (1986): *Evaluation. A systematic approach*. Beverly Hills: Sage.
- RUTMAN, L. (1980): *Planning useful evaluations. Evaluability assessment*. Beverly Hills: Sage.
- SCHALOCH, R. L. & THORNTON, C. V. (1988): *Program evaluation. A field guide for administrators*. New York: Plenum Press.
- SCHMIDT, R. E., SCANLON, J. W. & BELL, J. B. (1979): *Evaluability assessment: making public programs work better*. Rockville, Md.: Project Share, Department of Health and Human Services.
- SHIPMAN, S. (1989): General criteria for evaluating social programs. *Evaluation Practice*, 10 (1), 20-26.
- SHOM, E. D. (1978): *Glosary of evaluation terms*. Geneva: Joint Inspection Unit.
- SILVERMAN, M., RICCI, E. M. & GUNTER, M. J. (1990): Strategies for increasing the rigor of qualitative methods in evaluation of health care programs. *Evaluation Review*, 14 (1), 57-74.
- STANLEY, T. D. & ROBINSON, A. (1990): Sifting statistical significance from the artifact of regression-discontinuity design. *Evaluation Review*, 14 (2), 166-181.
- TRICKER, R. & DAVIS, L. G. (1988): Implementing drug education in schools: An analysis of the costs and teacher perceptions. *Journal of School Health*, 58, 181-185.
- VENEY, J. E. & KALUZNY, A. D. (1984): *Evaluation and decision making for health services programs*. Englewood Cliffs, N. J.: Prentice-Hall.
- WEISS, C.H. (1987): Where politics and evaluation research meet. In D. J. Palumbo (Ed.) *The politics of program evaluation* (pp. 47-70). Beverly Hills: Sage.
- WHEELER, G. R. & HISSONG, R. V. (1988): A survival time analysis of criminal sanctions for misdemeanor offenders: A case for alternatives to incarceration. *Evaluation Review*, 12 (5), 510-527.
- WHOLEY, J. S. (1983): *Evaluation and effective public management* Boston: Little & Brown.
- WILLIAMS, J. E. (1989): A numerically developed taxonomy of evaluation theory and practice. *Evaluation Review*, 13 (1), 18-31.
- WILLIAMS, W. (1975): Implementation analysis and assessment. *Policy Analysis*, 1 (3), 531-566.
- WRIGHT, D. J. G. (1989): A framework for dialogue on development policies and programs. *Evaluation Review*, 13 (4), 420-429.
- YORK, R. O. (1982): *Human service planning: Concepts, tools and methods*. Chapel Hill: University of North Carolina Press.