
Objetivos y destinatarios de las evaluaciones

Eric Monnier*

Este texto pretende ofrecer a los comanditarios, así como a las personas encargadas de la evaluación, los elementos necesarios para efectuar las opciones estratégicas para una evaluación. Esta no es una tarea fácil dado que numerosas problemáticas distintas, cuando no opuestas, se consideran propias de la evaluación como demuestran los abundantes calificativos utilizados particularmente en la literatura anglosajona para referirse a aquélla: *formative evaluation, summative evaluation, process evaluation, contextual evaluation, responsive evaluation*, etc.

¿De qué sirve una evaluación? Las críticas más duras de los comanditarios norteamericanos se refieren precisamente a la pobre utilización de los resultados de las evaluaciones. Pero esta cuestión nos lleva a plantear otra: ¿Cuáles son los objetivos de la evaluación proyectada? Evaluación se construye epistemológicamente a partir de «valor». Esta filiación semántica es la responsable de un contrasentido frecuente y se encuentra a menudo en el

origen de un *quid pro quo* entre, por una parte, los que concibieron la evaluación y, por otra parte, los que la encargaron.

El examen atento de los términos de la polémica metodológica revela que la cuestión clave, en cuanto a la utilidad de una evaluación, no es tanto la metodología como la credibilidad que la evaluación pueda tener para los actores sociales. Credibilidad que no depende, como se había creído, del rigor técnico de los instrumentos utilizados, sino de la legitimidad de la propuesta para los destinatarios de la evaluación. La evaluación no es nunca una práctica neutral y ajena a las relaciones de poder; es un hecho político que forma parte integrante del contexto político del programa, incluso si los «evaluadores»¹ intentan evitarlo y no quieren que sea así. Es por ello que proponemos en este texto una aproximación pluralista que asocie los actores sociales al desarrollo de la evaluación con el propósito de incrementar su credibilidad y, en consecuencia, su utilidad social.

* Laboratoire CEOPS Ecole National des Travaux Publics de l'Etat

1. Las Diferentes Aproximaciones a la Evaluación

Tyler, uno de los precursores en materia de evaluación, tuvo el mérito de proponer las bases de una aproximación que adopta como punto de partida los objetivos del programa considerado. Desde entonces, esta aproximación por los objetivos ha sido abundantemente criticada y otras aproximaciones han sido aplicadas. Sin ánimo de hacer una lista exhaustiva, examinaremos sucesivamente en qué se diferencian de la primera las otras cuatro grandes familias —aproximación por los efectos, por la forma en que el programa se puso en práctica, por los procesos y por los actores sociales.

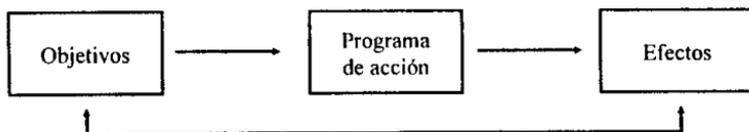
Para evitar un desarrollo excesivamente largo, nos referiremos a cada una de ellas a partir del caso más común de las evaluaciones *a posteriori*, dejando para el lector la interpretación de lo que correspondería a las evaluaciones *a priori* o concomitantes. Por otra parte, esta presentación nos brinda la oca-

sión de definir algunos conceptos claves de las evaluaciones.

La aproximación por los objetivos oficiales: excesivas suposiciones

La concepción más tradicional, probablemente la más frustrante, de la evaluación se centra en una cuestión muy simple: *¿En qué medida ha producido el programa los efectos esperados?; dicho de otra forma, ¿están los efectos observados en conformidad con los objetivos previstos?*

Esta aproximación, que confronta los resultados con los objetivos iniciales del programa, es aún utilizada muy frecuentemente en la actualidad. En efecto, un buen número de decisores consideran erróneamente que este tipo de evaluación, denominado *evaluación de impactos o de resultados*, es satisfactoria en la mayor parte de los casos. Esta aproximación se basa enteramente en la hipótesis que existen objetivos «claros, precisos y medibles» con los que se pueden relacionar los efectos, considerados asimismo medibles:



Si bien la mayor parte de las evaluaciones de este tipo no se limitan siempre a una contabilidad de los impactos (ni a un análisis de la relación costo/beneficios) sí que se encuentran con la dificultad de traducir en términos operacionales, racionales y jerarquizados los objetivos que son en general ambigüos incoherentes y hasta contradictorios. Por esta razón, muchas técnicas proponen procedimientos formales con el objeto de posibilitar que el programa sea «evaluable» según esta aproximación². Pero estas técnicas, entrevistas Delphi o análisis multicriterios, alcanzan, como máximo, a crear un consenso artificial (entre los diferentes actores sobre una definición común de los objetivos del programa) que se quiebra ineludiblemente en el momento de discutir los resultados de la evaluación³.

Por otra parte, esta aproximación se fundamenta en algunas otras construcciones, poco conformes con los procesos decisionales más comunes. Sabemos, por ejemplo, que la puesta en práctica de un programa diverge inevitablemente, en mayor o menor medida, del dispositivo inicialmente previsto; hay que poner en duda, por tanto, la atribución de efectos a la causa que originariamente se supone es el programa previsto. Así, innovaciones han sido rápidamente descartadas como

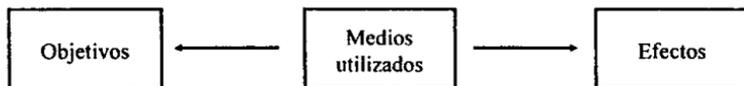
consecuencia de una evaluación de resultados cuando las organizaciones experimentaron dificultades durante el desarrollo del programa que no eran imputables a la innovación propiamente⁴. A pesar de estas consideraciones, una cuarta parte de las evaluaciones realizadas para el Departamento de Sanidad americano en 1970 no comportaban el menor análisis de la utilización efectiva de los medios previstos inicialmente. En otro terreno, una metaevaluación (una evaluación de 213 evaluaciones) de programas locales de ahorro de energía puesta en práctica en California puso de manifiesto que la mayor parte de los informes finales de los estudios no incluían descripción alguna, ni tan sólo elemental de los medios utilizados en el marco del programa evaluado⁵.

Aproximación por la puesta en práctica del programa: una perspectiva operacional

Una aproximación más realista consiste, pues, en considerar al programa no tanto como a una entidad homogénea invariante sino como a un conjunto de medios empleados (financieros, técnicos, de comunicación...). La cuestión central se formula esta vez de la siguiente forma: *La opción y la utilización de los medios, ¿ha permitido alcanzar los resultados esperados?* sien-

do la cuestión subsidiaria: *¿fueron empleados los medios de acuerdo con el proyecto y los objetivos iniciales?*

Y este enfoque se puede representar de acuerdo con el esquema siguiente:

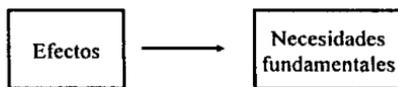


Estas *evaluaciones operacionales (implementation evaluation)* toman en consideración, además de la medida de la eficacia global del dispositivo, una maximización de los medios (realmente empleados) en relación con los fines (los objetivos fijados). Los experimentos más elaborados revelan precisamente una aproximación de este tipo. Las evaluaciones responden así más adecuadamente a su primera finalidad —la descripción de los hechos ocurridos durante el programa—, pero no alcanzan, sin embargo, a explicarlos. Las intenciones de los organizadores del programa ocupan como en el caso anterior, un lugar central que tiende, de una parte, a restringir la unidad de análisis a los efectos previsibles, y de la otra, a reducir la interpretación a los factores endógenos al programa.

Aproximación por los efectos: eliminar la influencia de los objetivos oficiales en las percepciones

Consciente de sus límites, M. Scriven

propuso en un texto que ha generado numerosas polémicas una aproximación que no tiene en cuenta para nada los objetivos oficiales: *las evaluaciones libres de objetivos (Goal-free evaluation)*⁶. Según este autor, las evaluaciones no deben centrarse en los objetivos iniciales, sino que deben más bien responder a la cuestión: *¿Cuáles son entre el conjunto de los efectos de un programa aquéllos que satisfacen las necesidades fundamentales?* Lo que se puede representar así:



Por iniciativa del Ministerio de Educación americano, M. Scriven había sido llamado a escoger entre los productos y métodos de enseñanza nuevos, aquellos que merecían ser difundidos. Esta actividad le permitió observar cómo muchos proyectos fracasan con respecto a sus objetivos iniciales, pero tienen en cambio consecuencias positivas en otros terrenos y son, por tanto, «recomendables». Un nuevo método de enseñanza de matemáticas, por ejemplo, puede no tener ningún impacto en la adquisición de estos conocimientos por parte de los alumnos pero incrementar, en cambio, su creatividad o incluso la comunicación en el seno de la clase.

Dicho de otra forma, poco importa que los objetivos iniciales no sean alcanzados si el programa produce efectos beneficiosos (aunque no intencionados) o, al revés, que los objetivos sean alcanzados al precio de efectos no deseados. Puede ocurrir, por otra parte, que diferentes actores sociales buscando conseguir efectos similares no estén de acuerdo sobre los medios a utilizar, o al contrario, que estando de acuerdo sobre la forma del programa esperen alcanzar efectos opuestos.

Una ilustración de este último caso la constituyen algunos programas de ahorro de energía que han producido lo

que se ha venido en llamar el «efecto boomerang» cuando se trataba de hogares de baja condición económica y particularmente de personas de edad avanzada: éstos incrementaron su consumo de calefacción inmediatamente después de las obras de rehabilitación de sus viviendas, reduciendo a la nada los ahorros de energía esperados. Para los responsables preocupados por el ahorro de energía, el programa fue, pues, un fracaso; pero los servicios sociales, preocupados por la salud de las personas mayores y los mismos beneficiarios, estaban en cambio convencidos de que la iniciativa había sido un éxito⁷.

El evaluador debe pues, según Scriven, guardarse de leer los folletos o de hablar con los responsables oficiales del programa antes de proceder al análisis de los efectos con el objeto de eliminar el sesgo que pueda derivar de un conocimiento previo de las expectativas. Esta precaución permite evaluar una «visión-túnel» y tener en cuenta en cambio, las connotaciones negativas derivadas de los efectos inesperados (cf. los calificativos habituales: efectos secundarios, periféricos, perversos...).

Pero esta aproximación ha sido poco utilizada puesto que su aplicación es delicada. En particular, para la deter-

minación de las famosas «necesidades fundamentales» el evaluador corre el riesgo de substituir con su propio sistema de valores el de los decisores⁸. Sin embargo, esta aproximación ha contribuido a profundizar en la reflexión sobre los *a priori* de toda evaluación.

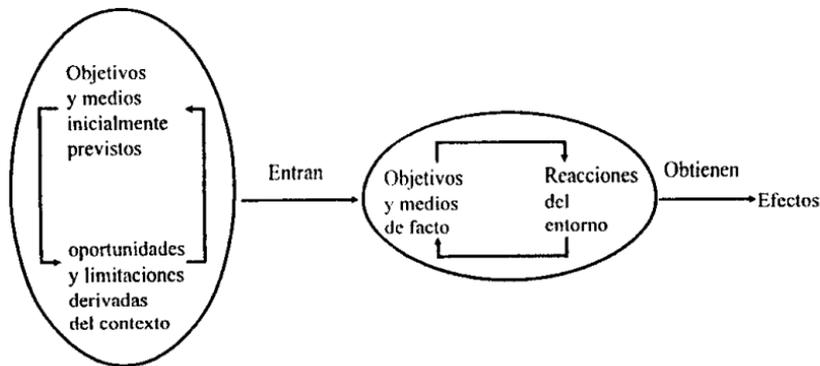
Aproximación por los procesos: las interacciones del programa y de su entorno

Las teorías sistemáticas inspiraron una aproximación por los procesos que tienen por finalidad principal explicar los fenómenos a partir de un análisis de las interacciones entre el programa y su entorno. Dicho de otra manera, esta vez la cuestión es: *¿Cuáles son los procesos que se encuentran en el origen de los efectos observables?*

Este tipo de *evaluación de procesos*

(«process evaluation») tiene, pues, por objetivo la identificación y la comprensión exhaustiva de los hechos relativos a las acciones emprendidas (a veces reciben la denominación «*comprehensive evaluation*»).

El programa y su organización son percibidos como entidades dinámicas que cambian a medida que se adaptan al contexto en el que son llevados a la práctica. Los objetivos iniciales pueden ser substituidos por nuevos objetivos operacionales, su jerarquía puede verse invertida y, por fin, los medios empleados *de facto* tienen su origen en un compromiso y divergen de los medios previstos *in abstracto*. La evaluación persigue, pues, reajustar los medios («una optimización») en función de las características del entorno. La idea puede ser esquematizada así:



En el contexto de esta aproximación, las intenciones de los responsables del programa mantienen todavía una cierta posición privilegiada, pero esta vez sus objetivos (y sus medios) son asumidos *de facto* y no en abstracto como ocurría con las evaluaciones operacionales mencionadas más arriba. Estos objetivos y medios *de facto* pasan a ser las referencias con las que se relacionan los efectos.

Diversos modelos de conducta para este tipo de evaluación han sido propuestos por los universitarios norteamericanos en respuesta a las nuevas exigencias de los decisores⁹. Son muy útiles en el caso de programas simples no excesivamente innovadores y de gran tamaño (desarrollados en varios sitios a la vez). En los demás casos, las evaluaciones de proceso plantean aún numerosos problemas prácticos mal resueltos (abundancia de variables independientes o explicativas) y continúan siendo muy caras, especialmente si se quiere mantener el paradigma experimental.

Por otra parte, los modelos propuestos conservan, por un lado, una visión sinóptica de los procesos decisionales, y por el otro lado, una aproximación en términos de sistema cerrado que persigue la exhaustividad. Las posiciones implícitas y los conflictos entre las di-

ferentes partes participantes son, pues, ignorados y la cuestión de la diversidad de los sistemas de valores de los actores afectados es dejada de lado.

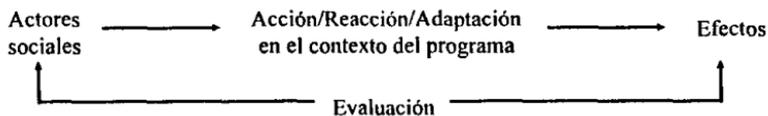
O bien los actores, cada vez más en número, ponen en duda, de una parte, la existencia de un consenso en torno a los objetivos (o el sistema de valores) y, de la otra, la posibilidad para una evaluación de alcanzar conclusiones objetivas e imparciales¹⁰.

Aproximación por los actores sociales: el reconocimiento del pluralismo

Esta última aproximación se construye a partir de los conflictos sociales emergentes más que no de los objetivos explícitos. Su punto de partida es el reconocimiento de una pluralidad de sistemas de valores que coexisten en el seno de la sociedad. El programa es percibido como un proceso nunca acabado de aprendizaje colectivo y de búsqueda pluralista de soluciones a los problemas sociales. La *evaluación pluralista*¹¹ participa, pues, de esta visión heurística y persigue responder a la siguiente cuestión: *¿Cuáles son las informaciones que requiere el sistema social para comprender mejor los efectos o hechos sobrevenidos a lo largo del desarrollo de un programa?*

La primera tarea de la evaluación consiste, por tanto, en identificar a los decisores y a los que utilizaron los resultados del análisis, siendo la segunda explicitar los sistemas de valores que orientarán los objetivos de la evaluación y sus características (método, cri-

terios...). El conjunto se concibe, pues, como un sistema abierto en el que la evaluación se integra, produciéndose una evolución de los objetivos de esta última en la medida que va desarrollándose.



Sistema evolutivo en el tiempo y en el espacio social

Cuadro I: La proposición de un nuevo enfoque

«Con el propósito de poner el acento en la evaluación de las cuestiones claves (*issues*) que son importantes para cada programa en particular, recomiendo el enfoque interactivo (*responsive approach*). Es una forma de abordar las cosas que pierde algo de precisión en los números con el propósito de incrementar su utilidad para las personas que son activas en el contexto del programa o que tienen un interés en éste. La evaluación de un método de educación es pluralista si examina las actividades reales más que los resultados perseguidos, si aporta una respuesta a las demandas de información de las personas interesadas y si la valoración de los fracasos y de los éxitos se hace teniendo en cuenta los diferentes sistemas de valores presentes.»

El enfoque propuesto implica, por tanto, realizar una serie de actividades, cada una de ellas más de una vez a lo largo de una evaluación y en un orden inspirado al evaluador por el curso de los mismos hechos.

1. Definición por el mismo evaluador de los límites del campo, de lo que es considerado como parte integrante del programa. Se basa para ello en los documentos oficiales, en los archivos del programa...
2. El evaluador observa directamente *lo que se hace* a título del programa y corrige las primeras ideas que se había hecho del programa.
3. Se encuentra entonces en situación de determinar a las partes que participan en el programa, sus *expectativas* con respecto al programa y sus objetivos implícitos y con respecto a la evaluación.
4. Se construye una *representación del conjunto de las cuestiones* y de los problemas que debe estudiar.

5. Se encuentra en situación de establecer una estrategia de evaluación y de seleccionar una determinada concepción (diseño), los datos y las informaciones que le harán falta recoger.
6. El evaluador selecciona su modo de recolección de datos y de informaciones en función de sus necesidades en cada situación. Generalmente, recurrirá al *juicio* de observadores humanos.
7. Puesta en marcha de la *recogida de datos*.
8. Presentación de los resultados: el evaluador prepara los instrumentos necesarios para conseguir una comunicación lo más natural que pueda, dando cuenta de forma extensa de experiencias personales.
9. El evaluador debe seleccionar las informaciones que difundirá de forma más amplia, determinar a quién hace partícipe de qué resultados, y *adaptar la forma al destinatario*.
10. Pone fin a la evaluación recogiendo una serie de elementos *de carácter conclusivo* y que se desprende de su trabajo de acuerdo con lo formalizado en su contrato de evaluador.

Fuente: Stake R. E., Gjerde C., 1974, *An evaluation of T-City. the twin city institute for talented youths*. Chicago, Rand McNally (Monograph Series in Curriculum Evaluation, núm. 7).

De acuerdo con esta aproximación, los actores no deben buscar una solución óptima sino una solución «satisfactoria» en función de un contexto dado. El sistema de valores y de referencias de cada actor evoluciona según encuentren satisfacción o se frustren sus expectativas, éstas últimas (el nivel de aspiración dirían los psicólogos) cambian en función de sus experiencias individuales y colectivas¹². Sucede con los objetivos lo mismo que con el horizonte, se alejan en la medida que se avanza. Cuanto más incierto es el horizonte, menos estables y concretos son los objetivos; dicho de otra forma, el futuro es menos previsible y más útil en la evaluación dado que reduce la in-

certidumbre de los decisores con respecto a su entorno.

2. La Utilidad Social de la Evaluación

Las dimensiones políticas de la evaluación

Podemos recordar sin duda cómo los científicos positivistas esperaban que la ciencia aboliese la política y que el gobierno de los hombres pasara a ser una rama de la estadística aplicada. Pero la realidad es otra: la evaluación no ha hecho caduca a la política, que se encuentra en el corazón de toda

evaluación. M. Q. Patton realizó una encuesta con los comanditarios gestores y evaluadores respectivos de veinte programas experimentales del Ministerio Federal de Sanidad que pone de manifiesto una imagen poco conocida de las condiciones en las cuales se realizan las evaluaciones¹³. Muchos de los interlocutores de M. Q. Patton admitieron que las consideraciones políticas habían influido en los juicios emitidos en el contexto de la evaluación, y también en la forma de trabajar. En efecto, la política no se reduce al campo de las representaciones o ideas expuestas por los políticos. En una sociedad, cualquier situación de arbitraje entre concepciones que no llegán a un consenso es política. Por esta razón, la introducción de un programa de acción de los poderes públicos es un proceso político, y de la misma manera lo es la evaluación: las opciones entre sistemas de clasificación, de referencias, los conceptos, las teorías, los modelos o hasta los puntos de vista adoptados o rechazados participan todos en los procesos políticos.

El carácter político de la evaluación se percibe también, sin embargo, en las condiciones sociales de *acceso a la información* que origina y en la toma de decisiones que genera. Por último, se percibe en el rechazo por parte del «evaluador» del mantenimiento de al-

gunas cuestiones formuladas por los organizadores o los destinatarios de un programa. No es nuestra intención discutir aquí el carácter beneficioso o nefasto de esta dimensión política, sino subrayar que la eventual utilidad de la evaluación no se puede dissociar de su vinculación con el campo político. Intentar huir de esta realidad representaría, en la mayoría de los casos, renunciar a toda utilidad social de la evaluación.

Estas observaciones vienen confirmadas por la constatación empírica efectuada por nuestros colegas norteamericanos a propósito de una serie de evaluaciones cuyos autores raramente habían anticipado que les conducirían a participar en algunos procesos políticos. De ahí se puede sacar una consecuencia que conecta directamente con las 95 tesis de CRONBACH y de sus «asociados»: «Una teoría de la evaluación tiene que ser tanto una *teoría de la interacción política* como una teoría relativa a la descripción de un hecho» (tesis núm. 11)¹⁴. Si se quiere definir con conocimiento de causa una evaluación es conveniente, por tanto, concebirla apoyándose en un análisis de la dinámica política en la cual se integra. Obviamente este análisis no se puede verificar durante la evaluación, pero su explicitación significa el principio de una

toma en consideración del carácter relativo de cualquier evaluación y, por tanto, de la posibilidad de modificaciones ulteriores¹⁵. En efecto, hemos observado anteriormente que la información y la comunicación son esenciales para alimentar las negociaciones que constituyen la base de cualquier programa público.

Después de insistir sobre las dimensiones propiamente políticas de las evaluaciones, conviene recordar que la evaluación no puede substituir a la decisión política. Parece incluso que cuando los responsables de las reformas sociales intentaron utilizar la evaluación de programas experimentales para favorecer una decisión, alcanzaron el resultado inverso. Esta es la constatación que cabe derivar del análisis de D. Greenberg y P. Robins a partir de una serie de experimentos sociales realizados en los Estados Unidos a partir de 1968¹⁶.

En efecto, entre 1968 y 1985 los programas de evaluación responden a la necesidad de justificar importantes reformas en el campo de la provisión de servicios sociales por el Gobierno federal. Estos programas han dado lugar a numerosos debates entre los políticos y los expertos. Estos últimos situaron sistemáticamente el debate en el cam-

po de los métodos y no consiguieron hacer prevalecer una sola manera de proceder. Los políticos, que tienen por costumbre, a falta de informaciones creíbles desde su punto de vista, recurrir a los principios éticos o «sociológicos», se vieron abocados ante la nueva situación a tomar posiciones de prudencia y expectantes. Por esta razón, las evaluaciones de los grandes programas sociales, y que desembocaron en las polémicas más encarnizadas, fueron la causa (como por error) de que se redujeran las posibilidades de que ciertas innovaciones fueran adoptadas por los decisores políticos.

Parece, pues, que las evaluaciones tradicionales fueron contestadas, independientemente del rigor metodológico, si no se preocupaban de encontrar una *legitimidad política*. Se había creído durante mucho tiempo que la independencia institucional y financiera del evaluador ofrecía una garantía de ola y, por tanto, una legitimidad (científico-técnica) suficiente. No fue así, sino todo lo contrario: únicamente situándose en el centro mismo de los procesos políticos, el evaluador evita el rechazo de las premisas en las que se basan las conclusiones. Una evaluación pluralista, sobre la base de un análisis colectivo de los datos, puede, a fin de evitar las dificultades derivadas de los virajes

políticos, reivindicar una «no-parcialidad» (noción que substituye a la de objetividad).

Llegados a este punto en la definición de las evaluaciones, es preciso señalar que antes que nada hay que preguntarse con qué finalidad se decide realizar una evaluación, además de definir desde qué puntos de vista se quiere llevar a cabo la evaluación de un programa y, por fin, explicitar los medios empleados y los efectos que se esperan de la acción. Por otra parte, es preciso darse una explicación tan satisfactoria como sea posible de los procesos políticos que constituyen el contexto en el que el programa y la evaluación tendrán lugar, con el objeto de precisar el lugar que ocupará aquélla.

Al formular estas recomendaciones, nos contentamos con plantear ambas cuestiones que pueden tener alguna utilidad práctica para la definición de una evaluación, pero no entramos, por coherencia con el contenido de los análisis procedentes, a proponer un modelo normativo que responda a todas y cada una de estas cuestiones.

Utilización de las conclusiones de la evaluación: las condiciones favorables

Hicieron falta muchas desilusiones por parte de los investigadores ante la poca

atención otorgada por los responsables políticos o por los gestores a sus resultados (generalmente muy caros), e igualmente muchas desilusiones de los responsables de los programas experimentales ante la naturaleza de las conclusiones de las evaluaciones, para que la cuestión de la utilización de los resultados se planteara de forma clara. La respuesta parecía evidente: una evaluación puede contemplar la valoración de las consecuencias sociales de un programa público desde diferentes puntos de vista. La interpretación de las relaciones de causa-efecto (entre los medios empleados en el marco del programa y las consecuencias posteriores) es, como se pudo constatar con los grandes experimentos sociales americanos, difícil y cara. Para formular las hipótesis relativas a los efectos producidos por un programa, lo que denominamos una teoría de acción, es preciso adoptar un determinado punto de vista. Y este punto de vista, ¿no es siempre el del poder público que aporta la financiación del programa? Ciertamente, el comanditario de la evaluación es a menudo la institución que financia el programa o que tiene la tutela administrativa de su puesta en práctica, pero ello no basta para definir un punto de vista, puesto que el aparato del Estado no es una entidad homogénea: distintos puntos de vista coexisten en su seno correspondiendo a cada

uno de los actores de los distintos niveles de un mismo organismo o a organismos que responden a misiones y estrategias institucionales variadas.

En este sentido, es preciso señalar que la evaluación puede ser objeto de utilidades perversas. Los autores americanos han propuesto una tipología de las perversiones que puede encerrar una demanda de evaluación: «evaluaciones-coartada», que consisten en mostrar las operaciones que mejor han funcionado para que sirvan de pantalla al análisis sistemático de un programa; «evaluaciones-aviso», en las que se señala un indicador de fracaso para justificar, sin más, la interrupción de un programa organizado contra la voluntad de la Administración bajo la presión de políticos locales o nacionales... Encontramos asimismo frecuentemente el caso de «evaluaciones-alego-prodomo», encargadas por la misma organización evaluada para justificar su existencia con respecto a otras instituciones (mantenimiento de un organismo, de una escuela...).

El responsable financiero de un programa adopta generalmente uno, al menos, de los tres puntos de vista siguientes:

1. ¿Cuál es la pertinencia y utilización efectiva de los medios empleados en el

contexto de la acción definida por el programa? (se corresponde con el punto de vista adoptado, por ejemplo, por las inspecciones generales de los ministerios franceses).

2. ¿Existe alguna posibilidad de mejorar la prestación de los servicios distribuidos en el marco del programa? (se corresponde con el trabajo habitual de las empresas de asesoría en organización).

3. ¿Qué nuevas medidas cabe tomar para resolver los problemas sociales que el programa ha permitido conocer mejor? (el Ministerio de Sanidad puede por ejemplo esperar que un programa experimental le sirva para descubrir en qué condiciones jóvenes drogados aceptarán una cura de desintoxicación).

Pero la perspectiva adoptada por los poderes públicos puede, asimismo, coincidir con la de uno de los actores del programa, excluyendo por tanto los puntos de vista de los demás (o de la misma Administración). El Ministerio de Sanidad en Gran Bretaña ha realizado en este sentido evaluaciones de programas de mantenimiento de enfermos graves en sus propios domicilios desde la perspectiva del miembro de la familia que se ocupa de los cuidados esenciales del enfermo. Esta perspectiva le ha llevado a modificar las directivas

impartidas a los médicos y a los profesionales que intervienen en estas situaciones.

Observamos, por tanto, como por la forma en que se define la evaluación y las cuestiones evaluativas hay tantas perspectivas posibles para, la interpretación de los datos y de las conclusiones de la evaluación como *situaciones socialmente definidas*. Podemos igualmente adoptar un punto de vista crítico ajeno a la forma como funciona realmente la sociedad. Es lo que hicieron muchos investigadores norteamericanos deseosos de fomentar un gobierno científico de las sociedades, dado que ellos creían que la resolución de los problemas sociales pasaba por la organización de experiencias *in situ*, de acuerdo con los cánones de la ciencia experimental. Esta visión puede hacernos sonreír hoy, pero conviene preguntarse cuáles fueron los efectos sociales de las evaluaciones practicadas de la forma como han sido practicadas generalmente hasta nuestros días (conforme al paradigma experimental).

La encuesta de M. Q. Patton citada aporta información muy interesante para llegar a comprender los factores favorables a una utilización real de los resultados de una evaluación por las distintas partes afectadas por un mismo programa. Para empezar, si bien la

utilidad de una evaluación es, a veces, fuertemente sentida por los actores, nunca se puede demostrar. En efecto, esta utilidad parece conseguirse en mayor medida en tanto en cuanto el proceso de evaluación suscita un mayor número de contactos entre, por una parte, el responsable del programa, su gestor, y por otra parte, el encargado de la evaluación: la evaluación aparece, pues, como un proceso que facilita y alimenta la reflexión, aunque nunca de una forma aislada. La evaluación se mezcla con las reflexiones que imponen el trabajo ordinario y las operaciones diarias. Por otra parte, es más eficaz en cuanto es más interactiva. Así es como la eficacia de la evaluación, desde la perspectiva de los decisores, aparece muy relacionada con su implicación en el proceso evaluativo. La evaluación, a pesar de los grandes proyectos de sus más ilustres teóricos, aparece, pues, más modestamente como un *proceso de transformación de las ideas y de los conocimientos* que contribuye a una evolución lenta o a transformaciones marginales de los programas sobre los que se aplica (esto en el mejor de los casos, aquellos en que los actores reconocen que ha jugado un rol nada despreciable).

Así pues, la utilidad social de una evaluación depende de las *condiciones de aprobación de sus propias conclusio-*

nes por los actores sociales a los que está destinada. Por ello defendemos que las evaluaciones pluralistas son las mejores cuando se trata de crear las condiciones para que se produzca una apropiación real y, por tanto, se utilicen sus resultados. Basándose en el diagnóstico de las distintas partes afectadas, en las cuestiones reales presentes y en los hechos relevantes, la aproximación pluralista recomienda suscitar un proceso *interactivo e iterativo*. Interactivo, en el sentido que la relación estrecha de los actores con el proceso de interpretación de los datos favorece la expresión de nuevas cuestiones, la aparición de nuevos actores y el replanteamiento de la forma de enfocar el problema social que se encuentra en el origen del programa evaluado. Iterativo, dado que la participación de los actores conduce a redefinir las bases del trabajo de evaluación.

Este tipo de evaluación se desarrolla, pues, de acuerdo con un proceso progresivo que debe converger en un compromiso sobre la interpretación de los hechos entre observadores y observador, de una parte, y entre los diferentes protagonistas del programa, por otra parte. Este proceso es, a nuestro entender, la base en la que se puede fundamentar la legitimidad de una evaluación. Una evaluación es creíble, y por tanto, puede ser útil socialmente

si es legítima ante los ojos de los actores sociales que tienen la capacidad necesaria para introducir las mejoras sugeridas por la misma evaluación.

Existe una evidente analogía entre el modelo de las evaluaciones pluralistas y el de los Círculos de Calidad en la industria. Los industriales aprendieron, bajo la presión de la competencia internacional (y bajo la influencia del modelo japonés, que substituyó al modelo sueco), las virtudes de la aproximación humana a los problemas organizacionales¹⁷. Al adoptar el principio de los Círculos de Calidad, los industriales llevaron a cabo una «revolución cultural»: renunciaron implícitamente al postulado tradicional según el cual existe una, y solamente una, mejor solución técnica y convinieron en que es preferible que las soluciones, aunque sean mediocres para los expertos técnicos, sean propuestas por aquellos que tendrán que aplicarlas. La participación de los destinatarios de la innovación en su concepción facilita su adopción y, en consecuencia, reduce las dificultades posteriores de su aplicación. De la misma forma, las evaluaciones pluralistas renuncian a la afirmación de conclusiones «rigurosas» (verificadas científicamente) para negociar, con los actores presentes, un compromiso aceptable para aquellos que deban sacar partido de las conclu-

siones para su actividad. La pérdida de rigor (o de calidad científica) es el precio a pagar para evitar el rechazo de la evaluación entera por sus destinatarios.

3. El Dispositivo de Evaluación

Evaluación endoformativa y evaluación recapitulativa: la toma en consideración de las expectativas de los destinatarios

Una de las distinciones más sutiles hechas en relación con el proceso de evaluación y que caló hondo en el otro lado del Atlántico fue propuesta por M. Scriven¹⁸; si la evaluación tiene por objeto informar a los protagonistas del programa para que puedan modificar sus conductas, mejorar su actividad y transformar así sus propios objetivos, diremos que la evaluación es *endoformativa* («*formative evaluation*»); si la evaluación está destinada a permitir que personas ajenas al programa —poderes públicos, políticos, la población en general— puedan formarse una opinión global acerca del valor intrínseco de la actividad, independientemente de la que puedan tener sus protagonistas, denominaremos a esta evaluación *recapitulativa* («*summative evaluation*»). Las dos situaciones pueden coincidir parcialmente, pero, general-

mente, son muy distintas. Para precisar un poco más, podemos añadir que la evaluación endoformativa viene a proporcionar una ayuda operacional a los organizadores del programa, mientras que la evaluación recapitulativa sirve el mismo tiempo a los que financian el programa desde fuera, para decidir sobre su lanzamiento, mantenimiento o reproducción, y otros actores sociales para juzgar la oportunidad de adoptar ellos mismos el programa evaluado.

Nuestros colegas Callon y otros llegaron a las mismas conclusiones. No obstante, utilizan una terminología muy distinta —evaluación interna y evaluación externa— que quizá se presta a confusión. Esta terminología se utiliza generalmente en referencia no a los destinatarios, sino a la pertenencia del equipo evaluador a la organización responsable del programa evaluado. De ahí que vayamos a utilizar los términos evaluación endoformativa y evaluación recapitulativa, dado que ambas pueden ser llevadas a cabo sin tener en cuenta si el equipo de evaluación pertenece o es ajeno a la institución responsable del programa.

Además, en el caso de la evaluación recapitulativa, los actores no implicados en el programa pueden desconfiar si la evaluación se hace de forma interna. De hecho, un equipo puede ser in-

terno a la organización de diversas formas. Así, por ejemplo, la mayoría de las grandes organizaciones internacionales, como el Banco Mundial o el Banco Interamericano, tiene en sus organizaciones tanto a un equipo de evaluación (*Review and Evaluation Office*) conectado con la Dirección General como diversos equipos de evaluación operacional (*Operations Evaluation Office*) en el seno de cada departamento sectorial.

Como hemos visto anteriormente, la concepción de una evaluación puramente recapitulativa era la gran ambición de los grandes experimentos de principios de los años 1970 en los Estados Unidos. El carácter esencialmente estéril de estos experimentos sociales, demasiado centrados en poner de manifiesto el valor «intrínseco» de una innovación, fue la causa del desarrollo del lado endoformativo de la evaluación.

El contenido de los dos tipos de evaluación difiere sensiblemente según el objetivo del programa innovador evaluado. Consideremos la distinción entre los programas cuyo objetivo es poner en marcha un prototipo de innovación y aquéllos que se dedican a adaptar un programa innovador a un contexto local distinto (los programas de transferencia de innovación). En el

caso de los programas locales de ahorro de energía en Francia, conocidos como «villes-pilotes»¹⁹, la agencia, en un primer momento, había seleccionado tres lugares donde fueron concebidos y probados diversos medios. Se optó por una evaluación endoformativa durante el curso del programa y por una evaluación recapitulativa para juzgar la oportunidad de reproducir este tipo de acción. Cuando, más tarde, la agencia decidió organizar una nueva serie de «villes-pilotes» los lugares fueron, intencionadamente, distintos de los primeros. Como no se había previsto en este caso una evaluación endoformativa, los resultados fueron más aleatorios. Uno de los lugares escogidos para implantar estos programas era de carácter eminentemente rural: se corrió el riesgo del fracaso más absoluto dado que los agricultores, destinatarios del programa, juzgaron que era de escándalo el lujo de una campaña de promoción que, sin embargo, había sido bien acogida en ciudades de tipo medio.

Dicho de otra forma, la transferencia de un programa nuevo plantea algunos interrogantes sobre el valor de un prototipo en un determinado contexto. Desde una perspectiva práctica, se trata de situaciones muy distintas que es preciso tener en cuenta. En el caso que no se haga un trabajo de construcción

sistemática de los objetivos de la evaluación, cabe esperar que se enjuicie una acción en función del sistema de valores de un actor local específico y que ello entrañe su rechazo dado que estos valores no serán reconocidos en otro contexto.

En el cuadro 2 se resumen las principales características de las evaluaciones que pueden resultar del cruce de una finalidad endoformativa o recapitulativa de la evaluación y de una situación de puesta a punto de un prototipo o de transferencia de un programa nuevo, explicando a su vez los objetivos de la evaluación y sus destinatarios, así como las fuentes de referencia utilizadas. Esta cuestión es algo difícil de precisar, dado que no existe un método para resolver la dificultad que en-

traña, pero es importante recordar que forma parte integrante del enunciado de la situación.

Esta discusión puede ciertamente parecer algo abstracta, pero es importante no cometer el error de creer que no tiene ninguna importancia práctica. Al contrario, al preguntarse, antes de iniciar un programa nuevo eventualmente muy costoso, lo que debe ser su evaluación, se pueden reducir gastos considerables y evitar disgustos innecesarios.

La evaluación pluralista: un lugar de negociaciones en cascada

En su forma más tradicional, la evaluación de una acción pública se lleva a cabo según un procedimiento que

Cuadro 2: Características de una evaluación según sus finalidades

Tipo de evaluación	Programa que persigue	
	La puesta a punto de un prototipo de programa nuevo	La transferencia de un programa nuevo
Endoformativa	<i>Objetivos:</i> concepción y mejora de la innovación. <i>Destinatarios:</i> los protagonistas de la innovación	<i>Objetivos:</i> adaptar una innovación a un contexto. <i>Destinatarios:</i> el equipo local encargado de la innovación.
Recapitulativa	<i>Objetivos:</i> validar o garantizar la innovación. <i>Destinatarios:</i> responsables nacionales, colegas, expertos, profesionales, usuarios potenciales	<i>Objetivos:</i> certificar o delimitar el campo de aplicación de la innovación. <i>Destinatarios:</i> responsables locales interesados en la innovación, responsables nacionales.

consta de cuatro fases: de entrada, la Administración tutelante, generalmente aquella que decidió en su día financiar el programa, pide una evaluación a un equipo independiente. En una segunda fase, éste se entrevista con los responsables de la intervención y mide los efectos. La tercera fase consiste en la redacción de un informe final de evaluación. Esta constituye la más importante para los «evaluadores» tradicionales. Por último, en cuarto lugar, el informe es difundido y desata la agitación política a no ser que sea prudentemente guardado dentro de un cajón para no volver a salir a la luz pública.

Este procedimiento, que se adapta perfectamente al modelo burocrático y tecnocrático, debe ser claramente denunciado; es contraproducente, puesto que el precio que hay que pagar por él es considerable. Sabemos en efecto, por experiencia, que la recopilación de datos es más difícil, más costosa y a veces hasta imposible dentro de la función pública sin la cooperación de los «evaluados». De otra parte, el riesgo de contrasentido o de interpretación superficial de los datos es proporcional a la resistencia de las partes interesadas en la evaluación. En fin, acabamos de verlo, la utilidad social de los resultados de las evaluaciones depende estrechamente de la relación de confian-

za y de colaboración entre analistas y analizados.

Estas correlaciones, bien conocidas por los psicoanalistas, fueron constatadas en el marco de la metaevaluación de Freeman y Bernstein²⁰. Según estos autores, en efecto, la calidad de una evaluación está correlacionada positivamente con el grado de interdependencia entre el programa y la evaluación. Dicho de otra forma, estos autores constataron empíricamente que, en el conjunto de evaluaciones estudiadas «las investigaciones, en las que los responsables del programa y los de la evaluación pertenecían a la misma organización, eran de mejor calidad que las producidas por los evaluadores de una organización independiente del programa».

Las evaluaciones realizadas completamente desde fuera carecen del conocimiento cualitativo esencial sobre lo que ha pasado realmente y alimentan un conflicto crónico entre los evaluadores y los responsables directos del programa evaluado. Estas consideraciones nos llevan a cuestionar las recomendaciones de la Comisión del Comisariado del Plan en el sentido de adoptar la regla deontológica llamada de la «separabilidad», según la cual la calidad de una evaluación (podemos traducir por «objetividad») depende de

la independencia financiera e institucional del equipo encargado con respecto al organismo evaluado.

¿Hay que concluir, por tanto, en sentido inverso que las evaluaciones deberían ser siempre llevadas a cabo en el seno del mismo programa? Esta no es nuestra opinión, puesto que si ponemos el acento en la necesaria relación de confianza entre la persona encargada de la evaluación de una parte y el conjunto de los protagonistas de la acción de la otra, una gran proximidad entre uno, y uno sólo de entre ellos, supone un riesgo de comprometer la credibilidad de su trabajo con respecto a los demás protagonistas. Como, por otra parte, una persona raramente puede pertenecer a varias organizaciones, su independencia institucional parece recomendable a fin de evitar la colusión con uno de los protagonistas en detrimento de los demás.

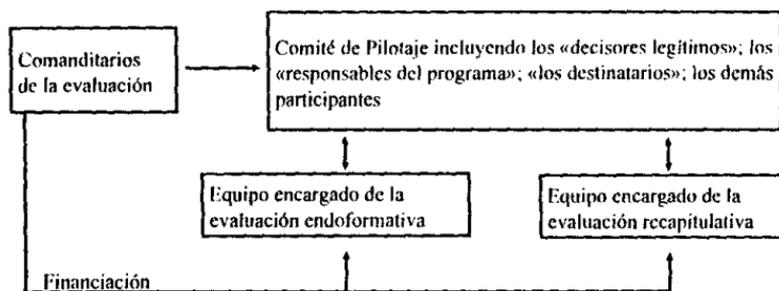
Para que una evaluación constituya verdaderamente un *centro de negociación* se puede ciertamente contar con la calificación profesional de la persona encargada de la evaluación, pero nosotros recomendamos ante todo poner en funcionamiento un dispositivo idóneo (cf. cuadro 3).

El elemento esencial del dispositivo propuesto es claramente lo que hemos

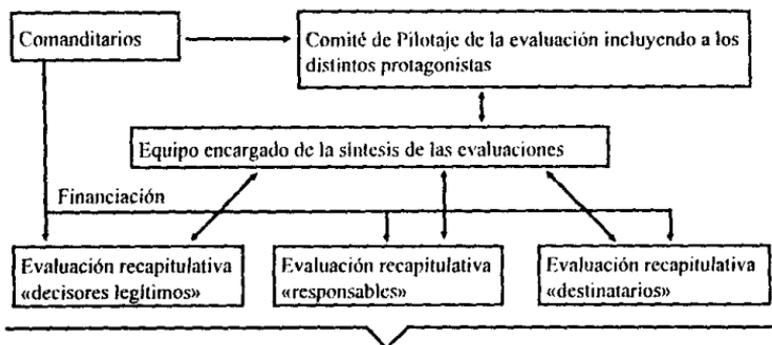
llamado un «comité de pilotaje», que agrupe a los diversos protagonistas, que seleccione al equipo encargado de la evaluación y por tanto de todos los trabajos. Con un dispositivo de este tipo, los comandatarios no tienen mayor poder que los demás miembros del comité sobre la orientación de la evaluación además de concienciarse de la progresión de la reflexión colectiva. De esta forma, el informe final de la evaluación pasa a ser un documento puramente formal destinado únicamente a las personas externas (redactado con el propósito de ajustarse a las cláusulas contractuales), que describe los instrumentos de trabajo utilizados, las informaciones recogidas y pasa cuentas del compromiso negociado entre los protagonistas a propósito de la interpretación de los hechos y de las mejoras posibles que cabe introducir con respecto al nuevo programa.

Hay que precisar aquí que nosotros distinguimos la negociación de la concertación tal y como es practicada habitualmente, puesto que no es cuestión de alcanzar una perfecta representatividad de las diferentes partes en el seno del comité de pilotaje, ni de creer en la igualdad de sus pesos respectivos en el marco de las negociaciones.

En el caso que una evaluación endo-

Cuadro 3. Ejemplos de dispositivos para las evaluaciones pluralistas

En el caso que una evaluación endoformativa no pueda ser realizada



El que realiza la evaluación adopta la perspectiva del protagonista por el que ha sido seleccionado

formativa no pueda ser llevada a cabo, parece razonable realizar la evaluación recapitulativa sobre la base de varias evaluaciones paralelas confiadas a equipos seleccionados por cada uno de los protagonistas del programa. Si la financiación no permite un dispositivo de este tipo, serán suficientes dos evaluaciones, una que haga de abogado

del programa y la otra que se oponga a éste²¹.

La evaluación es (debe ser concebida como) un proceso pluralista pero también evolutivo y polifinalista. Si no, acostumbra a convertirse en un procedimiento rígido, completamente orientado hacia una finalidad única de veri-

ficación del buen respeto de las reglas administrativas o de la buena gestión de los medios empleados.

La formulación de las finalidades de la negociación varía según que la evaluación tenga lugar antes, durante o después de la acción. En el caso de una evaluación endoformativa podemos contar al menos *siete finalidades* de las negociaciones²²:

1. Poner en práctica un sistema integrado de recogida y tratamiento de las informaciones (económicas, sociales y operacionales) gracias a la colaboración activa de los participantes en la intervención.
2. Mostrar las limitaciones (o dificultades encontradas) de orden técnico, económico, cultural o político, relacionadas con el entorno en el que la innovación se introdujo.
3. Conceptualizar nuevos medios (o ajustes operacionales) reconocidos por los diferentes actores como susceptibles de aportar una solución a un problema social considerado.
4. Clarificar la naturaleza del problema social que motivó la reacción original.
5. Poner de relieve conflictos en la in-

terpretación de los hechos y definir los objetivos de la nación.

6. Aprender de una terminología y de los conceptos comunes por la aculturación recíproca de los distintos protagonistas; dicho de otra forma, la constitución y la generalización de lo que se ha convenido en llamar un «referente» común.

7. Estimular el cambio social a partir de una confrontación constructiva de los valores y de las expectativas de las que son portadores los diversos componentes de la sociedad, gracias a la puesta en común de capacidad creativa.

Esta lista no es ni exhaustiva ni jerarquizada, sólo la presentamos aquí a título indicativo para señalar la importancia del campo cubierto por las negociaciones y el interés para un comanditario y para un encargado de evaluación de reconocer su papel.

Notas

¹ El término «evaluador» se refiere al equipo o al conjunto de equipos coordinados encargados de la realización de la evaluación. Nosotros utilizaremos el término «analista», ya que nos parece corresponde mejor al espíritu de nuestras recomendaciones.

² Entre los más conocidos citamos: Cook T., y Campbell, D. (1979), *Quasi-experimentation: design and analysis issues for field settings*. Chicago, Rand McNally; y Rossi, P. H., y Freeman, H. E. (1982), *Evaluation: a systematic approach*, Londres, Sage.

³ Patton, M. Q. (1978), *Utilization focused evaluation*, Londres, Sage. En esta obra estas técnicas: estrategia de la «pirueta», de la formulación de objetivos contradictorios o de la «guerra de objetivos».

⁴ Freeman, H. E. (1977), «The present status of evaluation research», en *Evaluation Studies Review Annual*, vol. 2.

⁵ White, L. T.; Archer, D.; Aronson, E. et al. (1984): «Energy consumption research of California's utilities», en *Evaluation Review*, vol. 8, núm. 2.

⁶ Scriven, M. (1973), *Goal Free Evaluation*, en House, E. (ed.); *School Evaluation: the Politics and Process*, New York, McCutchan.

⁷ Monnier, E. et al. (eds.) (1986), *Consumer Behaviour and energy policy: an international perspective*, Londres, Preager.

⁸ Patton, M. Q., 1978, *op. cit.*

⁹ Cf. los modelos citados en las siguientes obras:

— Etzioni, A. (1968), *The Active Society: a theory of societal and political processus*, Free Pres.

— Perrow, C. (1970), *Organizational Analysis:*

a sociological view. Wadsworth.

— Stufflebeam, D. L. et al. (1971), *Educational Evaluation and Decision-Making*, Peacock Publisher.

— Monnier, E. et al., *Evaluation Methods for Energy Conservation Program in the Residential Sector*, en Monnier, E. et al. (eds.), *Consumer Behaviour and Energy Policy: an international perspective*, New York, Preager.

¹⁰ A pesar de las diferencias en algunos casos importantes en los puntos de vista de sus propuestas, existen muchos puntos en común entre autores como Cronbach, Guba, Lincoln, y Patton. Cf:

— Cronbach, L. J. et al. (1980), *Toward reform of Program evaluation*. San Francisco, Jossey Bass.

— Guba, E.; Lincoln, Y. (1981), *Effective Evaluation*, San Francisco, Jossey Bass.

— Patton, M. Q. (1978), *op. cit.*

Cf: también las diferentes contribuciones sobre el tema: Weijs, C. (ed.) (1983) «Stakeholder-Based Evaluation» en *New Directions for Program Evaluation*, núm. 17, San Francisco; Jossey Bass.

¹¹ Encontramos en la literatura americana numerosas denominaciones de lo que llamamos evaluación pluralista:

— *Responsive Evaluation*, citado en Stake, R. E.; Gjerde, C. (1974), *An Evaluation of T. City, The twin city Institute for talented youths*. Chicago, Rand McNally (Monograph Series in Curriculum Evaluation núm. 7).

— *Active-reactive-adaptive evaluation*, citado en Patton, M. Q. (1978), *op. cit.*

¹² Patton, M. Q. (1978), *op. cit.*

¹³ Patton, M. Q. (1978), *op. cit.*

¹⁴ Cronbach, L. *et al.* (1980), *op. cit.*

¹⁵ El estudio de la pertinencia de estas teorías sobre el funcionamiento político de ciertas organizaciones proviene evidentemente de la investigación en ciencias sociales y no de las evaluaciones. Esto significa que estas últimas pueden sacar mucho provecho de apoyarse en las primeras.

¹⁶ Greenberg, D.; Robins, P. (1985), «The changing role of social experiments in policy analysis», en *Evaluation Studies Annual Review*, vol. 10.

¹⁷ Midler, C. (1986), «Logique de la mode managériale», en *Annales des Mines*, junio.

¹⁸ Scriven, M. (1967), *The methodology of Evaluation*, en Tyler, R.; Cagne, R.; Scriven, M. (eds.), *Perspective of Curriculum Evaluation*, Chicago, Rand McNally (Monograph Series in Curriculum Evaluation, núm. 1).

¹⁹ Monnier, E. (1985), «Maitrise de l'énergie dans le secteur domestique», en *Cahiers du CSTB*, jan-février, núm. 1972.

²⁰ Citado por Brewer, M.; Struening, E. (1983), *Handbook of Evaluation Research*, Londres, Sage.

²¹ Este principio de investigaciones contradictorias fue sugerido por Rossi, citado en Cronbach, L. J. (1980), *op. cit.*

²² Esta enumeración ha sido adaptada a partir de las conclusiones penetrantes de L. Nizard a propósito de las «funciones» de la elaboración del sexto Plan. Nizard, L. (1974), *Changement social et appareil d'Etat: du 5ème al 6ème Plan*, St. martin d'Hères: CERAT.

Fuente

Revista Documentación Administrativa No. 224-225, octubre 1990-marzo 1991.

Traducción: Xavier Ballart