

¿Cómo evaluar programas y servicios públicos?

Aproximación sistemática y estudios de caso

Xavier Ballart

Ministerio para la
Administraciones Públicas

1992

Colección: Estudios

Serie: Administración del estado

ISBN: 84-7088-623-1

Este material se utiliza con fines
exclusivamente didácticos

CAPÍTULO III. CONCEPTOS, CLASES, PROPÓSITOS, INTERESES Y CRITERIOS DE EVALUACIÓN

1. CONCEPTO

La evaluación de programas no tiene una definición pacífica y ampliamente aceptada por la literatura. Los términos «program evaluation», «policy evaluation» y «evaluation research» han sido objeto de una variedad de interpretaciones en función de lo que distintos investigadores incluyen en este campo de investigación y/o entienden que son sus métodos propios.

A. AMPLIACIÓN PROGRESIVA DEL AMBITO DE LA EVALUACIÓN

Inicialmente, la evaluación de programas es percibida como la evaluación de los efectos netos de los programas –es decir, de aquellos efectos que no se hubieran producido en ausencia del programa– en relación con los objetivos inicialmente propuestos. (Rossi y Wright, 1984:51).

En este contexto, la evaluación es, pues, un análisis de resultados («outcomes» en inglés), entendiendo a éstos como el efecto o impacto real de un programa en sus destinatarios.

Interesa aquí precisar la diferencia que existe entre «outcomes» y outputs, puesto que a menudo ambos se confunden. Los outputs se refieren a lo que el programa produce. Así, por ejemplo, un programa de rehabilitación de viviendas produce viviendas aptas para ser ocupadas por ciudadanos. Los «outcomes», por su parte, se refieren al impacto de estos outputs en la sociedad. En el ejemplo propuesto, el «outcome» del programa sería el efecto que produce dicha política de rehabilitación en la condición de personas de bajo nivel de renta y, en concreto, en su capacidad de acceder a una vivienda de calidad.

Esta concepción de la evaluación de programas como evaluación de los efectos de un programa en relación con sus objetivos es recogida en el influyente libro de Weiss (1972:4):

«El propósito de la Investigación Evaluativa es medir los efectos de un programa en relación con los objetivos que se propone alcanzar, como medio de contribuir al proceso subsiguiente de toma de decisiones sobre el programa, y por tanto como medio de mejorar la programación futura».¹

La evaluación de los efectos netos de los programas trataba a éstos como a una unidad indiferenciada, lo que se ha convenido en llamar una caja negra, sobre la que no se conoce nada sobre sus componentes.

Así, por ejemplo, un programa de formación e inserción profesional puede estar compuesto de clases teóricas sobre donde encontrar ofertas de trabajo, cómo hacer un currículum o presentarse en una entrevista de selección, puede incluir módulos de calificación profesional, de prácticas en una empresa u otros. La evaluación de sus efectos trataría al programa como a una unidad indiferenciada comparando, por ejemplo, los niveles de éxito de aquéllos que habían participado en el programa con los de aquéllos que no lo habían hecho y formaban parte de un grupo de control. Esta evaluación aportaría una estimación objetiva de los efectos netos del programa, pero dado que no entraría en el análisis de sus componentes, no aportaría información sobre el porqué había o no había funcionado.

De aquí que varios autores incluyan en sus definiciones de evaluación de programas, la evaluación de los componentes del programa y de la forma en que el programa opera regularmente. Esta concepción ha dado lugar a las «management-centered evaluations» defendidas por Wholey:

«Evaluación 1. valora la eficacia de un programa en funcionamiento, en alcanzar sus objetivos, 2. se basa en los principios del diseño investigativo para distinguir los efectos de un programa de aquéllos de otras fuerzas

¹ En el mismo sentido, Walker que estudió la función evaluativa en la GAO piensa que: «el objetivo de la evaluación es determinar si los programas federales están obteniendo los beneficios y resultados perseguidos por el Congreso y la Administración» (1985-173).

También, Nioche en Europa, define la evaluación en parecidos términos: «utilizar los métodos de investigación de las ciencias sociales para apreciar los efectos reales de una política, una vez que ésta ha sido aplicada» 1982b: 21).

operando en una situación determinada, y 3. se dirige a mejorar el programa a través de la modificación de las operaciones corrientes.» (Wholey et al., 1975:23)².

Se relaciona, por tanto, con la preocupación por hacer de la evaluación un instrumento que resulte útil a los responsables directos de los programas. En este sentido, Subirats (1989:140) se refiere a una perspectiva gerencial que distingue de la perspectiva legal-tradicional y de otras de carácter político³.

La evaluación aún ampliaría su ámbito a raíz de los análisis de evaluaciones realizadas en los que se sugería que los pobres resultados de muchos programas podían ser debidos a una inadecuada planificación previa, a un diagnóstico erróneo de las necesidades de sus destinatarios o a problemas de interpretación o puesta en práctica de conceptos manejados en las fases de discusión, negociación y aprobación del programa (Rossi y Wright, 1984:65).

De otra parte, como se ha afirmado anteriormente, la crisis industrial de los años setenta y las políticas conservadoras de reducción del gasto público suscitaron un mayor interés por la evaluación económica de los programas, lo que ha llevado a algunos autores a incluir dentro del ámbito de la evaluación de programas los análisis coste-eficacia y coste-beneficio desarrollados por economistas. (Lewis y Zarb, 1974:308; Thompson, 1980: entero).

B. CONCEPTO AMPLIO DE EVALUACIÓN DE PROGRAMAS

Todo ello conduce a que algunos eminentes científicos sociales como Rossi y Freeman empezaran a hablar de evaluaciones globales («comprehensive evaluations») como prototipo de las actividades de investigación evaluativa. Para estos autores la evaluación de programas consiste en:

«La aplicación sistemática de procedimientos de investigación social para valorar la conceptualización y diseño, implementación y utilidad, de los programas de intervención social» (1985:19-20).

Esta definición resume los puntos indicados anteriormente y establece las bases de una tipología de estudios de evaluación.

En primer lugar, la evaluación es aplicación sistemática de procedimientos de investigación social. Existe, por tanto, un compromiso por seguir las prescripciones y los métodos de las ciencias sociales con el fin de producir evidencia empírica válida y fiable.

En segundo lugar, la evaluación se aplica con el fin de valorar y eventualmente perfeccionar, primero, la *conceptualización y diseño*, segundo, la *implementación*, y tercero, la *utilidad* de los programas de intervención social. La evaluación completa de un programa incluiría, por tanto, estas tres cuestiones. Rossi y Freeman utilizan el término genérico de «evaluación de utilidad» para referirse a la evaluación de *impacto o efectos netos* y a la evaluación *económica o de eficiencia*.

En tercer lugar, el ámbito de la investigación evaluativa se limita a los programas de intervención social, esto es, a cualquier combinación de objetivos y recursos que se organiza para que tenga un efecto en el bienestar de individuos o grupos sociales concretos. Programas, por tanto, muy diversos que pueden tratar de educación, bienestar social, salud, vivienda, seguridad, rehabilitación social, desarrollo económico.

En resumen, tal y como indica la figura número 2, la evaluación de programas tiene que ver tanto con los efectos de un programa sobre sus beneficiarios, como con su implementación, como con el proceso de formulación de propósitos generales y objetivos más específicos, lo que tendría que permitir relacionar, de acuerdo con el modelo teórico, resultados finales, con la forma en que es administrado el programa y planificada la intervención.

² Para Marvin y Hedrick, analistas de la GAO, la evaluación de programas consiste en «la valoración sistemática de cómo un programa está alcanzando los objetivos establecidos en la legislación autorizante» (1974: 328).

En el mismo sentido, Herman, Morris y Fitz-Gibbon entienden que evaluar un programa consiste en: «recoger información, de forma sistemática, sobre cómo opera un programa y sobre los efectos que pueda tener.» (1987: 8).

³ Ciertamente, los criterios en función de los cuales valorar el éxito de una política o programa pueden ser plurales y hasta contradictorios. A ello nos referimos al tratar de valores, propósitos e intereses.

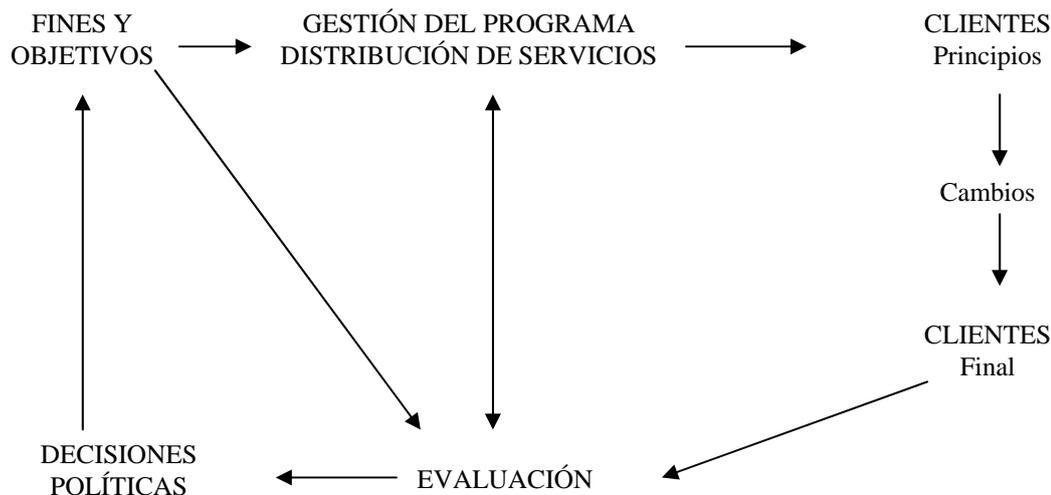


Figura 2: Modelo teórico para la evaluación de programas

C. DIFERENCIAS ENTRE EVALUACIÓN DE PROGRAMAS Y CONTROL ADMINISTRATIVO

Si bien, en algunos países, las instituciones tradicionales de control externo de la Administración han añadido a sus funciones de auditoría jurídico-contable las funciones de evaluación en los términos que acabamos de explicar, hay que distinguir la evaluación del control y de la inspección administrativa.

De una parte, porque las cuestiones a las que se dirigen son distintas: el control administrativo no cuestiona la oportunidad de una política o programa, en cuanto conjunto de recursos aplicados a la consecución de objetivos externos, sino que únicamente verifica que los recursos previstos hayan sido bien aplicados y que las normas existentes hayan sido respetadas. En el mismo sentido, Vulpian y Devez (1984:278).

De otra parte, porque los controles se ejercen desde una perspectiva organizacional y por tanto, estrechamente relacionada con cada una de las estructuras administrativas existentes. La evaluación, en cambio, privilegia la perspectiva longitudinal basada en la idea de política o programa. Rompe, por tanto, como afirma Nioche (1982a:51), con la problemática de tipo jerárquico e implica una aproximación a la vez más distanciada y más global, sobre políticas o programas en los que pueden intervenir varias administraciones.

Finalmente, porque como indica Eveno (1984:219), la inspección y el control administrativos se basan en los desplazamientos sobre el terreno, en las constataciones de hechos observables y en el juicio personal del inspector. El resultado de su trabajo se refleja generalmente en un informe escrito que tiene por destinatario el superior jerárquico. La evaluación es más amplia, es más exigente en cuanto al método e implica por tanto más tiempo.

2. CLASES DE EVALUACIÓN

Dado que cada evaluación suele tener un ámbito distinto en función de las cuestiones que intenta responder, resulta útil distinguir entre cuatro clases de evaluación que coinciden con las grandes cuestiones incluidas en la definición de Rossi y Freeman.

Lógicamente, la tipología de la evaluación se puede construir en función de otros criterios, lo que da lugar a otras tantas clases de evaluación. Así, en función del propósito específico que motiva la evaluación se distingue entre evaluación formativa o de proceso y evaluación conclusiva o de resultados; en función de la naturaleza del estudio, entre evaluación descriptiva y evaluación analítica; y en función de la relación de los evaluadores con la organización que encarga la evaluación, entre evaluación interna y externa. Finalmente, incluimos una breve referencia a la meta-evaluación.

A. CLASIFICACIÓN POR FASES

a. *Evaluación de la Conceptualización y del Diseño de los Programas*

La evaluación de la conceptualización y del diseño de los programas intenta proporcionar información, primero, sobre el problema del que se ocupa o se debería ocupar el programa, segundo, sobre el proceso de formulación y diseño del programa y tercero, sobre la racionalidad intrínseca del mismo programa, es decir, hasta qué punto el diseño del programa sirve sus objetivos (Rossi y Freeman, 1985:38-39).

Incluimos aquí, por tanto, la revisión de los estudios de necesidades donde se recogen datos sobre la incidencia, prevalencia, y distribución en el espacio físico y social del problema en cuestión, así como la investigación básica sobre los procesos políticos y sociales que tienen que ver con la formulación, negociación y aprobación de nuevas intervenciones (Rossi y Wright, 1984:65).

El estudio de caso sobre el Plan Integral de Roquetes es en buena medida una evaluación de esta naturaleza⁴.

b. *Evaluación de la Implementación de los Programas*

Este segundo tipo de evaluación consiste en la valoración sistemática y periódica de dos cuestiones fundamentales: la cobertura del programa, es decir, hasta qué punto el programa está alcanzando la población o el área a los que se dirige; y la forma en que el programa se administra, tanto desde el punto de vista de la utilización de los recursos de que dispone, como de los servicios finalmente distribuidos (Rossi y Freeman, 1985:40).

Se trata, por tanto, de una supervisión o control del funcionamiento del programa que indudablemente puede interesar a los gestores del programa que pueden tomar acciones de tipo correctivo en relación con problemas de gestión (problemas de distribución, de utilización de recursos humanos, de flujo de la información) pero que también puede interesar a aquéllos que deciden sobre su continuidad y su financiación y que por tanto tienen un interés evidente en conocer los problemas que existen con su puesta en práctica, independientemente de su preocupación por los resultados finales.

El estudio de caso sobre la Policía local de l'Hospitalet de Llobregat responde en buena medida a este tipo de evaluación.

c. *Evaluación de la Eficacia o Impacto de los Programas*

Una evaluación de eficacia o impacto examina la medida en que éste produce un cambio en la dirección deseada. Lógicamente, si no existe este cambio o impacto, o no es demostrable, resulta más difícil defender su continuación.

El análisis de impacto presupone que existe un conjunto de objetivos operacionales y que éstos son indicativos del éxito o del fracaso del programa. Se afirma que éste ha funcionado correctamente si consigue algún cambio hacia estos objetivos. La función del evaluador consiste precisamente en demostrar que los cambios son debidos a la intervención y no se pueden explicar como consecuencia de un proceso alternativo que no incluya al programa mismo (Rossi y Freeman, 1985:43).

La cuestión esencial en la evaluación de impacto es, pues, separar los efectos del programa de otros efectos externos para determinar hasta qué punto el programa es responsable de los cambios observados durante y después de su puesta en práctica. Esto se consigue con la identificación de medidas de resultados que representen los objetivos del programa, y con el control, mediante métodos experimentales, cuasiexperimentales o estadísticos, de aquellos efectos que no sean efectos reales del programa.

⁴ También existe una literatura sobre el concepto de problema. Generalmente se habla de problema cuando concurren dos elementos: por una parte la constatación de una situación no deseada en la condición humana y social, y por otra, la resolución decidida de tomar una acción organizada con el propósito de remediar aquella situación. La definición de problema implica por tanto la idea de acción dirigida a tratar el mismo problema.

En el mismo sentido, Subirats destaca el carácter «artificial» de los problemas, en la medida que «no tienen vida propia al margen de los individuos o grupos que los definen.» Por otra parte, vincula problema y solución. Por ello, el planteamiento de un problema se ha de hacer «desde una perspectiva que posibilite la intervención y que permita una vía de mejora» (1989: 50).

Tanto el estudio de caso sobre el Plan Integral de Roquetes como el de la Policía Local incluyen elementos que se aproximan al concepto de evaluación de eficacia.

d. Evaluación de la Eficiencia de los Programas

El conocimiento de la eficacia o impacto de un programa puede ser insuficiente. A menudo, el impacto de un programa deberá ser juzgado en función de su coste ya que puede ocurrir perfectamente que un determinado programa esté produciendo unos resultados escasos por los costes en que ha incurrido. Ello supone incluir el análisis coste-eficacia y el análisis coste-beneficio dentro del ámbito de la evaluación. Como se dijo anteriormente, sin embargo, éstos requieren una evaluación de impacto previa puesto que únicamente han de tenerse en cuenta los beneficios realmente causados por el programa. Por otra parte, la identificación, cuantificación y transformación a una unidad común de medida de los costes y beneficios puede resultar problemática cuando se trata de programas sociales, ya que no todos los outputs e inputs pueden ser valorados en términos monetarios.

Cuadro 2: Principales cuestiones para cada tipo de evaluación

1. Evaluación de la conceptualización y del diseño de los programas

- *¿Cuáles son el grado y la distribución de los problemas a los que se dirige el programa?*
- *¿Se ha diseñado el programa de acuerdo con los objetivos inicialmente perseguidos? ¿Es lógico y racional?*

2. Evaluación de la implementación de los programas

- *¿Hasta qué punto el programa está alcanzando la población específica o el área a los que se dirige?*
- *¿En qué medida el esfuerzo de intervención se está llevando a cabo como se especificó en el diseño del programa?*

3. Evaluación de la eficacia o impacto de los programas

- *¿Ha sido el programa efectivo en alcanzar sus objetivos iniciales?*
- *¿Se pueden explicar los resultados del programa como consecuencia de un proceso alternativo que no incluya al programa mismo?*

4. Evaluación de la eficiencia de los programas

- *¿Cuáles son los costes y los beneficios de distribuir los servicios a los participantes en el programa?*
- *¿Supone, el programa en concreto, un uso eficiente de los recursos en comparación con otros alternativos?*

Fuente: Adaptado a partir de Rossi y Freeman, 1985:39-46.

B. OTRAS CLASIFICACIONES

a. Evaluación Formativa - Evaluación Conclusiva

Los términos «formativa» y «conclusiva» (o sumativa, de «summative» en inglés) fueron adoptados por Scriven en su discusión sobre la evaluación de programas educativos (Scriven, 1967), y desde entonces ha sido utilizados con asiduidad por la literatura para referirse a evaluaciones con propósitos distintos. (Herman, Morris y Fitz-Gibbon, 1987:17).

La evaluación *formativa* responde a una preocupación que se encuentra en la base de muchas demandas de evaluación: obtener información sobre la marcha de un programa, de forma progresiva durante su desarrollo, con el objeto de darla a conocer a sus directores y personal, y poder así éstos introducir las modificaciones programáticas que se estimen oportunas.

Ello requiere un conocimiento detallado de las operaciones del programa y de sus múltiples componentes para poder anticipar problemas y examinar en qué medida el programa se está desarrollando conforme a sus objetivos.

La evaluación *formativa* o de proceso coincide, en cuanto a las cuestiones objeto de investigación, con la evaluación de la conceptualización, del diseño y de la implementación de los programas.

El propósito de la evaluación *conclusiva o de resultados* es proporcionar la información necesaria para poder hacer una valoración global sobre los resultados de un programa y sobre su utilidad.

Esta clase de evaluación requiere que haya pasado un intervalo suficiente de tiempo desde la adopción del programa de forma que su organización y actividades hayan alcanzado una cierta estabilidad. En la medida que el programa tenga unos objetivos claros y medibles será más factible juzgar sus resultados, comparándolos, por ejemplo, con los de programas similares o con los alcanzados por un grupo de control que no haya recibido ningún programa.

La evaluación *conclusiva* o de resultados se corresponde, en cuanto a las cuestiones objeto de investigación con las evaluaciones de impacto y de eficiencia.

La matriz de la figura 3 relaciona los propósitos que generalmente motivan la demanda de una evaluación con los aspectos del programa objeto de investigación.

Propósito	Cuestiones de mayor interés		
	Concept. Diseño	Implementación	Impacto/Coste
Formativa	X	X	(X)
Conclusiva			X

Figura 3: Tipos de evaluación de programas

b. Evaluación Descriptiva - Evaluación Analítica

El objeto de las evaluaciones descriptivas es obtener parámetros cuantitativos que puedan ser aplicados a grupos de individuos que se benefician de una política o participan en un programa y/o a las organizaciones que los administran.

En cambio, las evaluaciones analíticas intentan responder a cuestiones de tipo causa-efecto. Requieren por tanto la utilización de modelos teóricos y diseños de investigación adecuados para demostrar que las interpretaciones causales son ciertas. (Cohen y Franco, 1988:116).

En la medida que las evaluaciones de la implementación de los programas persiguen presentar información sobre la población objeto del programa y sobre las actividades en que se concreta aquél, suelen ser de naturaleza descriptiva. Mientras que las evaluaciones de impacto, en cuanto intentan separar los efectos del programa de otros efectos externos, suelen ser de naturaleza analítica.

c. Evaluación Interna - Evaluación Externa

La distinción entre evaluación interna y externa se refiere a las personas encargadas de realizar la evaluación y plantea un problema clásico de organización de la evaluación, esto es, ¿quién debe ser el evaluador: personal del programa, especialistas contratados por la Administración responsable del programa o especialistas externos completamente ajenos a aquélla?

Se puede definir una serie de criterios generales conforme a los cuales determinar las ventajas e inconvenientes que presentan la evaluación interna y externa.

En primer lugar, si la evaluación es realizada por el personal del programa, su comprensión del mismo y de los problemas que pueda plantear será máxima. También es cierto, sin embargo, que la evaluación podría fácilmente convertirse en una evaluación-justificación. En este sentido, parece importante confiar cualquier evaluación a personas cuya libertad de enjuiciar el programa evaluado esté garantizada.

La contratación de especialistas por la misma Administración responsable del programa puede servir para conseguir una visión más alejada y global de éste, así como una mayor competencia y por tanto una evaluación de mejor calidad. Los riesgos de falta de objetividad e independencia persisten, sin embargo, en la medida que satisfacer al cliente con unos buenos resultados puede ser la única vía de obtener un nuevo contrato.

El recurso a expertos externos ajenos al programa y a la Administración que lo gestiona presenta evidentes ventajas en relación con la objetividad, independencia y competencia de los evaluadores, pero también el riesgo de no poder llegar a conocer la realidad del programa, o incluso de producir unos resultados marginalmente útiles a la Administración evaluada por un problema de falta de confianza en sus propósitos. Para el personal del programa, la presencia de un evaluador externo podría únicamente responder a la necesidad de encontrar los datos que apoyen un juicio negativo, hecho «a priori». (Bocialetti y Kaplan, 1986:2)

En resumen, pues, el dilema está entre una evaluación interna que dispone de la información, pero carece del espíritu crítico y quizás de la competencia necesarios, y una evaluación externa, que puede tener la libertad de enjuiciamiento y competencia necesarias, pero encontrarse con dificultades para reunir la información pertinente.

Ciertamente, la evaluación se ve reforzada por los estudios externos y múltiples (Wildavsky, 1980:7), aunque es preciso conseguir la aceptación y la participación del personal del programa, tanto para desarrollar la evaluación misma, como para que sus resultados sean después utilizados. En este sentido, las fórmulas mixtas, bajo la dirección de expertos externos y dependencia directa de un organismo distinto del que administra el programa, siempre que sea posible, aparecen como las más apropiadas⁵.

d. Metaevaluación

La metaevaluación consiste en el análisis de los resultados de varias evaluaciones. Proporciona, por tanto, información sintética sobre el conocimiento existente en relación con una determinada política o programa público. Lógicamente, este tipo de análisis sólo es posible cuando un número considerable de evaluaciones han sido realizadas y publicadas en relación con una determinada política. Entonces, el examen sistemático del conjunto de las evaluaciones, desde una perspectiva más general, puede proporcionar más y mejor información que un nuevo estudio de características parecidas a los realizados anteriormente⁶.

A veces, la metaevaluación consiste en repetir un estudio acabado, reanalizando los datos en que se basa o incorporando más datos sobre el mismo programa. (Hudson y McRoberts, 1984:222). Otra forma de hacer metaevaluaciones consiste en la revisión de tipo más amplio de la investigación evaluativa desarrollada en un determinado período. El trabajo de Wholey et al (1975) cubrió 15 programas de cuatro agencias norteamericanas durante 1968-1969 y examinó tanto la organización de programas, como la calidad de las evaluaciones y de sus informes finales, y los problemas de comunicación y posterior uso de los resultados. Nioche y Poinard realizaron una investigación de naturaleza similar para Francia (1984:5-22).

La «evaluación de las evaluaciones,» como podría describirse la metaevaluación, no está exenta de las complicaciones que derivan de las diferencias existentes entre proyectos individuales en cuanto a las metodologías empleadas, el tipo de datos empleados, o la perspectiva política desde la que se llevaron a cabo. Parece, sin embargo, razonable recurrir a la información agregada y sintética de las metaevaluaciones para justificar decisiones políticas, cuando esto sea posible, y en la medida que la evidencia empírica acumulada indique claramente un curso de acción.

⁵ Para una discusión más completa de las ventajas e inconvenientes de la evaluación interna y externa, Vid. Cohen y Franco, (1988: 111-114); Monnier (1987: 116); o Weiss (1972: 20-21).

⁶ Por ejemplo, en el campo de la salud y medio ambiente, Urkowitz y Laessing (1982: 427-438) recogen un conjunto de investigaciones y examinan su rigor metodológico. Su principal objetivo es descubrir los efectos de los problemas de tipo medio ambiental en la salud humana. Pero para ello intentan determinar la credibilidad de los resultados de las investigaciones realizadas, al mismo tiempo que señalan aquellos aspectos que requieren más investigación y que por tanto deberían incluirse en las agendas de los investigadores.

3. PROPOSITOS

A. PROPOSITOS ABIERTOS

Como quedó reflejado en la figura 3, la evaluación de una política o programa puede abarcar un conjunto de cuestiones diversas en función de los propósitos o razones por los que se decida llevar a cabo. Igualmente hemos indicado dos propósitos o razones principales por las que en general se decide evaluar una política o un programa: de una parte, obtener información sobre un programa, de forma progresiva durante su desarrollo, con el objeto de determinar las modificaciones programáticas que precisa: y de otra parte, proporcionar la información necesaria para poder hacer una valoración global sobre los resultados de una política o programa y sobre su utilidad. Así, cabría esperar que una evaluación formativa fuera seguida de cambios en el personal, en las actividades, la organización o incluso en el diseño de un programa. Mientras que la información proporcionada por una evaluación de resultados finales tendría que servir de ayuda a los órganos político-administrativos de los que depende un programa para decidir sobre su continuidad o suspensión, sobre su generalización o su reducción⁷.

El principal propósito de la evaluación es por tanto producir información que tenga alguna relevancia para aquellas personas que encargan la evaluación y que se entiende han de tomar algunas decisiones en relación con la política o programa evaluado.

B. PROPOSITOS ENCUBIERTOS

Algunos autores se han referido, sin embargo, a los posibles propósitos encubiertos de la evaluación lo que nos lleva a introducir algunas consideraciones más políticas.

Ciertamente, una evaluación se puede hacer con el propósito de justificar una posición a la que se ha llegado de forma previa: ya sea ésta la de ofrecer una imagen positiva del programa, o al revés, poner de manifiesto sus resultados negativos con el objeto de utilizarlos como arma política en contra de sus responsables políticos y/o administrativos⁸.

Por ello parece importante, desde la perspectiva del evaluador, determinar los intereses y motivaciones que se encuentran detrás de la demanda de evaluación y, concretamente el rol que es llamado a representar, de forma consciente o supuestamente inadvertida.

La evaluación también se puede utilizar como un medio para posponer una decisión. La demanda de más investigación como requisito previo a emprender algún tipo de acción es estrategia común en el ámbito de las Administraciones Públicas. Ya nos hemos referido a como, desde mediados de los años setenta, la evaluación ha contribuido a frenar el desarrollo creciente de las políticas sociales en algunos países.

En otro sentido, probablemente muchos programas son evaluados, simplemente porque así lo requiere la legislación que establece su financiación (Levitan y Wurzburg, 1979:3). Lógicamente ello conduce a una evaluación de carácter ritual que se realiza únicamente para cumplir con las disposiciones legislativas pero que no tiene ninguna utilidad para los administradores de los programas.

Al mismo tiempo, no creemos que se pueda considerar ilícito buscar la legitimación que pueda aportar la evaluación para defender una posición, cuando en el seno de una Administración existe un debate sobre una nueva política o programa. Al contrario, este tipo de prácticas, salvo que la evaluación sea cuestionada me todo lógicamente, pueden contribuir a enriquecer el diálogo político.

⁷ En el mismo sentido, Chelimsky relaciona los propósitos de la evaluación de programas con la formulación de nuevas políticas o programas, con la gestión y administración de los ya existentes y con la exigencia social de dar cuentas de los resultados de los programas en funcionamiento y ya consolidados (1978: 8). Por otra parte, Weiss indica la necesidad de delimitar claramente el propósito de una evaluación, en la medida que esta autora entiende que la evaluación que sirve todos los propósitos («the all-purpose evaluation») es un mito (1972: 15).

⁸ En el primer sentido, Weiss se refiere a la evaluación como instrumento de auto-glorificación:

«El administrador cree que su programa funciona muy bien y busca la manera de hacerlo visible ... Copias del informe, favorable por supuesto, pueden ser enviadas a los órganos directores, a los miembros de las comisiones legislativas ... » (1972: 12).

4. INTERESES Y CRITERIOS

A. MULTIPLICIDAD DE INTERESES

Hasta ahora nos hemos referido a las personas o grupos de personas con capacidad para tomar decisiones sobre una política o programa. Las políticas y programas públicos, sin embargo, afectan a una pluralidad de grupos que van más allá de los decisores políticos-administrativos y del personal encargado de llevarlos a la práctica.

Así, si tomamos como ejemplo la reforma de la enseñanza secundaria en España, veremos que además de interesar a los Parlamentos centrales y autonómicos, al Ministerio de Educación y Ciencia y a los departamentos de enseñanza autonómicos, interesa a municipios, a asociaciones profesionales de enseñantes, sindicatos, investigadores universitarios, y por supuesto a profesores, alumnos, padres de alumnos y a sus asociaciones.

El carácter de afectado por una política o por un programa no tiene porque suscitar, sin embargo, un mismo interés en la evaluación del mismo. El tipo de información que querrían obtener cada uno de estos grupos varía en función de su posición en el sistema.

Así, en primer lugar, cabe esperar que aquéllos que tienen una mayor capacidad para decidir sobre el futuro de la reforma (parlamentarios, miembros del gobierno central y autonómico, y altos funcionarios) tengan un mayor interés en obtener información sobre resultados finales de los estudiantes en tests estandarizados, sobre las características de los estudiantes que se inscriban en cada una de las nuevas opciones o sobre sus dedicaciones futuras. Información que les ayude a responder cuestiones de carácter general sobre la conveniencia o no de continuar con la aplicación de la reforma, institucionalizar nuevos programas o mantenerlos como experiencias piloto, aumentar el esfuerzo presupuestario en la línea marcada por la reforma o limitarlo.

En segundo lugar, los directores y responsables de poner en marcha la reforma, tanto a nivel central, como autonómico y local, estarán más interesados en saber, no sólo si la reforma se desarrolla conforme a las previsiones, sino también qué aspectos generan mayores problemas, cuáles aparecen como más positivos o qué estrategias administrativas alcanzan mejores resultados de forma más eficiente.

En tercer lugar, los profesionales afectados y que trabajan en contacto con los beneficiarios teóricos de la nueva política educativa generalmente se preocuparán más de cuestiones relacionadas con sus nuevas condiciones laborales y con la forma de realizarse su trabajo diario. En este sentido son más pragmáticos y esperan obtener información que les facilite la adaptación al nuevo sistema y, en su caso, aprender aquello que deberían hacer y no están haciendo.

En cuarto lugar, corresponde a los alumnos, padres de alumnos y a sus asociaciones, adoptar la posición de cliente, y desde esta perspectiva, requerir la información necesaria para valorar la calidad de los servicios recibidos. Por otra parte, los ciudadanos en general, como contribuyentes, estarán preocupados por saber, aunque sea en términos muy generales, cuanto cuesta la reforma y más específicamente, en qué se gasta su dinero.

Es, pues, importante *reconocer la diversidad de intereses en juego y las múltiples perspectivas que estos intereses implican a la hora de enjuiciar una política*. También es cierto, sin embargo, que una misma evaluación difícilmente podría proporcionar información de inmediata utilidad para todos los grupos, dada la diversidad de sus preocupaciones.

B. ¿EVALUACIÓN, CONFORME A QUE CRITERIOS?

a. *Los Valores e Intereses de los Actores*

El reconocimiento de la existencia de una pluralidad de intereses en relación con una determinada política o programa nos lleva a plantear la cuestión de los criterios. A pesar de que la literatura de la evaluación ha prestado poca atención a este tema, dado que inicialmente se creyó que la actividad de la evaluación debía estar exenta de valores, es obvio que el mismo término «evaluación» lleva implícita la idea de juzgar el mérito, esto es, el valor o la bondad de las políticas y de los programas. Cabe preguntarse por tanto, ¿conforme a qué criterios?

El conocimiento de los intereses de todas las partes afectadas por el programa evaluado puede ser utilizado para seleccionar los criterios de evaluación, de forma que ningún criterio de especial importancia para un determinado grupo no sea omitido. La comprensión de todos los intereses presentes de acuerdo con

una orientación pluralista, puede ser importante para obtener la cooperación de todos los grupos o para que los resultados de una evaluación no sean descartados en base a acusaciones de partidismo por parte de los evaluadores.

La descripción de los intereses de cada grupo, sin embargo, si bien sirve al efecto de conocer las distintas perspectivas de valoración, tampoco genera los principios específicos en función de los cuales justificar la selección de algunos criterios por encima de los demás.

b. Los Objetivos del Programa

Tradicionalmente, se ha utilizado los objetivos de los programas como criterio de evaluación. Ello tiene sentido en la medida que aquéllos suelen ser el producto de la negociación política y, en este sentido, reflejan en buena parte los intereses que participan en el proceso de formulación y aprobación de las políticas. No obstante, la evaluación por objetivos no está exenta de problemas:

En primer lugar porque los objetivos no siempre están claros. Generalmente son abstractos, múltiples y hasta contradictorios.

En segundo lugar, porque los programas pueden tener efectos no esperados que pueden ser tan importantes como los buscados expresamente.

En tercer lugar, porque el personal de un programa tiende a fijar objetivos más bien modestos, en la medida que es mejor pasar un listón bajo que no estrellarse en uno excesivamente alto. En cambio, los responsables políticos de los programas tienden a hacer lo contrario, especialmente en las etapas iniciales de declaración política de intenciones o cuando es preciso conseguir financiación externa.

En cuarto lugar, porque los objetivos de los programas cambian en función del entorno local donde se desarrollan, de forma que una misma política aplicada en distintos lugares puede dar lugar a programas completamente diferentes⁹.

En la medida que la evaluación no recurra a los objetivos del propio programa, el valor o mérito de éste dependerá de la totalidad de los efectos o cambios producidos, sean éstos intencionados o no intencionados, positivos o negativos. (Cook y Shadish, 1986:49).

c. Los Objetivos Teóricos de la Intervención Pública

Algunos autores han intentado sistematizar los objetivos más importantes de los programas públicos, con el propósito de ofrecer una lista de criterios que sirvan para la evaluación de cualquier programa como inventario de referencia.

Así, para Langbein (1980), los principales *objetivos* de los programas públicos son el *crecimiento económico* y la *distribución de la riqueza*.

De una parte, el crecimiento económico se puede promover incrementando la productividad de los recursos existentes o aumentando la eficiencia de los mecanismos de mercado y eliminando sus «fallos» (1980:14). *Productividad y eficiencia* aparecen pues como dos criterios generales de crecimiento económico. El avance con respecto a uno de ellos no supone necesariamente progresos con respecto al otro en la medida que puede tratarse de objetivos contradictorios.

Así, como afirma la misma autora, un programa que incentiva la utilización de recursos improductivos (tierras marginales, trabajadores parados...) no necesariamente promueve el uso más productivo de estos recursos y puede de hecho redundar en una menor eficiencia económica del sistema en su conjunto, al restar medios a otras actividades más productivas. Sin embargo, es cierto que el incremento de la productividad de recursos no utilizados es uno de los objetivos fundamentales de una buena parte de los programas públicos, y como tal, se puede medir el grado en que se alcanza este objetivo (1980:15).

Por otra parte, el perfeccionamiento de los mecanismos de mercado se puede conseguir, según Langbein a través de:

- «1) la producción de bienes no privados;
- 2) ofreciendo mejor información a los agentes del mercado o reduciendo los costes de una información imperfecta;

⁹ Como se verá con el estudio de caso, éstas son las dificultades con las que nos encontrábamos a la hora de evaluar el Plan Integral de Roquetes.

- 3) distribuyendo el riesgo de inversiones arriesgadas;
- 4) promoviendo la participación en el mercado de todos los agentes y eliminando cualquier tipo de restricción o discriminación» (1980:15)

Así, para evaluar los efectos de la legislación que protege el medio ambiente, habría que determinar si realmente produce un bien público; para evaluar la reglamentación administrativa de las etiquetas de los productos nutritivos, habría que determinar la medida en que se ofrece mejor información al consumidor; etc.

De otra parte, el segundo gran *objetivo* de los programas públicos es, según Langbein, la *redistribución de la riqueza*. La evaluación de las políticas que persiguen este objetivo se puede realizar de acuerdo con los tres *criterios* siguientes: 1. *Nivel de beneficios que ofrecen los programas*, 2. *Equidad horizontal*, y 3. *Equidad vertical*. El primero se refiere a la medida en que los beneficios ofrecidos por el programa son suficientes o adecuados. El segundo se refiere a la medida en que el programa llega y trata de igual forma, a todas las personas potencialmente eligibles. Y el tercero se refiere al grado en que personas no incluidas en el grupo al que va dirigido el programa se benefician de éste.

Así, un programa de formación profesional, puede ser muy efectivo en conseguir un empleo para los jóvenes que lo terminan (lo que en sí constituye un primer criterio de evaluación), pero al mismo tiempo, puede ser muy injusto desde la perspectiva de equidad horizontal, en la medida que únicamente una pequeña proporción de jóvenes parados tuvieron acceso a él, o desde la perspectiva de la equidad vertical, en la medida que los que se beneficiaron realmente de él no se corresponden con su población-objetivo.

Un último criterio propuesto por Langbein es el de la *satisfacción de los ciudadanos*, «en la medida que en una sociedad democrática debería haber una cierta correspondencia entre los deseos de sus miembros y los programas de sus instituciones de gobierno.» (1980:22)

Seguramente, el inventario de objetivos y criterios que acabamos de repasar no es completo. Objetivos más políticos como los que encontramos en el caso del Plan Integral de Roquetes, –el incremento de la participación de los ciudadanos en las actividades de la Administración, la consecución de niveles más altos de coordinación entre los agentes públicos y privados que operan en un territorio o el perfeccionamiento de los mecanismos de respuesta a las necesidades locales por parte de la Administración–, encuentran difícil acomodo en el esquema de Langbein. Su aplicación obligaría, en muchos casos, a forzar la traducción de programas no económicos, en términos de crecimiento económico. La clarificación del mapa de objetivos y criterios, no obstante, habría de permitir al evaluador no dejarse ningún aspecto relevante por evaluar, aunque insistimos, su utilidad es más bien a nivel indicativo.

C. OTROS CRITERIOS ALTERNATIVOS

Otros autores, en cambio, han propuesto criterios alternativos, que nada tienen que ver con los objetivos de los programas, a partir de los cuales juzgar el valor de una política o programa y que consisten en:

1. examinar el grado en que se han satisfecho las necesidades materiales de la población o área objeto del programa, independientemente de los objetivos de éste (Scriven, 1973);
2. examinar aquello que los clientes, personal y administradores del programa entienden se ha conseguido con éste, así como aquellos factores que creen llevaron a su éxito o fracaso (Guba y Lincoln, 1981);
3. examinar los factores que tengan una mayor importancia para aquéllos que tengan que tomar decisiones sobre el programa en un futuro próximo (Patton, 1986);
4. examinar la medida en que se han satisfecho las necesidades de información de cada una de los grupos afectados por el programa (Bryk, 1983);

Cada una de estas propuestas representa, en mayor o menor medida, una aproximación alternativa al paradigma racional-científico que domina la evaluación. Interesa ahora entrar en la discusión de los principales modelos de evaluación, con el objeto de responder a una cuestión difícil: ¿Cómo evaluar las políticas y los programas públicos?

Ciertamente, no se trata de importar mecánicamente prácticas de evaluación desarrolladas mayoritariamente en los EEUU. Se trata, más bien, de sacar partido de los conocimientos y de la experiencia allí acumulados, para a partir de estudios de caso como los propuestos, extraer algunas implicaciones que puedan tener alguna relevancia para la evaluación, en general, y más específicamente, en nuestro entorno.