

Metaevaluación de políticas públicas: una visión desde la ciencia política

Ester García Sánchez

1. A vueltas con el carácter político de la evaluación¹

Las evaluaciones son procesos eminentemente políticos. Hasta hace unas décadas, sólo los autores próximos a enfoques “cualitativistas” -como MacDonald (1977), House (1980) o Guba y Lincoln (1981; 1987; 1989), entre otros- hubieran suscrito tal afirmación. En líneas generales, desde estos enfoques se solían esgrimir tres argumentos para explicar por qué las evaluaciones son escenarios especialmente impregnados de connotaciones políticas:

- Las evaluaciones tienen como objeto de estudio unos elementos -las políticas y los programas públicos- que son propuestos, elaborados, debatidos, financiados y aprobados *en el seno del proceso político* (Weiss, 1987; Palumbo, 1987). Su dinámica, sus estructuras y sus actores -en especial, aquellos implicados en la política pública evaluada- inciden notablemente en su realización.

- El impulso hacia su consolidación procede, fundamentalmente, de los poderes públicos, en forma de institucionalización y/o de financiación. Los gobiernos ven en ella un elemento de legitimación al tiempo que una oportunidad para el control. Al fin y al cabo, como afirma House (1993), el gobierno legitima la evaluación y la evaluación legitima la actuación gubernamental.

- Las evaluaciones se realizan con el propósito -tácito o explícito- de informar e influir, en un sentido o en otro, en el proceso de toma de decisiones políticas. No quiere ello decir que se orienten siempre a la mejora de la política o el programa en cuestión. Con frecuencia, se emplean como medio de rentabilizar el “esfuerzo político” o, incluso, se convierten en un arma arrojada entre opciones de distinto signo. Son pues especialmente proclives a ser instrumentalizadas y utilizadas para fines políticos o partidistas².

Pero no son estos los únicos argumentos que pueden aducirse. Como ya señalamos en otro lugar (García Sánchez, 2005), para profundizar en el tema resulta oportuno acudir al potencial analítico que puede ofrecernos la Ciencia Política. Esta disciplina y, en particular, los estudios sobre los actores, los procesos de toma de decisiones o los más recientes de *políticas públicas*³ resultan ser, para nuestros propósitos, una referencia inexcusable. Desde la Ciencia Política, la evaluación se concibe:

- Como una actividad intrínsecamente política desde el momento en que en ella se ven involucrados actores⁴ con intereses contrapuestos, políticamente relevantes, que luchan por obtener una mayor cuota de poder e influencia en el proceso evaluativo, con el propósito de modularlo de acuerdo a sus preferencias. Y es que las evaluaciones, lejos de constituir escenarios eminentemente racionales, asépticos e integrados en torno a unos objetivos comunes, están presididas por la diversidad de intereses. Los distintos actores tienen percepciones y aspiraciones distintas acerca de cómo debe ser la evaluación y del papel que en ella les debe corresponder. Cada actor lucha por hacer valer sus pretensiones y por obtener la mayor cuota posible de poder en sus transacciones. El escenario de la evaluación se define, así, a partir de la interacción política de los actores implicados y/o con intereses en la evaluación.

- Configurada a partir de las decisiones (o no decisiones) que adoptan los distintos actores en el transcurso del proceso evaluativo. Estas decisiones -por ejemplo, qué actores participan, de quién es la propiedad de la información producida o a qué enfoque metodológico se recurre- implican un determinado posicionamiento sobre la política pública que se evalúa y sobre el propio proceso de evaluación; en definitiva, sobre las asignaciones de valor que en ellos se realizan⁵. Coincidimos, pues, con Greene, cuando afirma que la evaluación siempre “defiende” o “promueve” unos determinados valores e intereses y descarta otros (Harvard Family Research Project, 2005).

Hoy en día prácticamente nadie cuestionaría la validez de estas afirmaciones, poniendo en duda que las evaluaciones tienen, además de una dimensión técnica -en tanto que procesos de investigación-, un marcado carácter político. La evaluación, señala Thomas Schwandt (2005: 5), “es una práctica tanto política como científica”. Pero es cierto también que, como nos recuerda con cierta sorpresa este mismo autor, aún hay quienes creen que la política representa una amenaza para la evaluación, el *lado de la subjetividad* frente a la dimensión técnico-científica en la que reinan los hechos objetivos y las explicaciones contrastadas, y que los evaluadores deberían intentar minimizar la influencia de la política en sus evaluaciones.

Sin embargo, por fortuna, cada vez son más los autores que reconocen que las evaluaciones son procesos inherentemente políticos, aunque entre ellos existan diferencias acerca de cómo se debe gestionar la “política de la evaluación” (Greene, 2003)⁶. Buena prueba de ello es que, en los últimos años, el interés por la relación política-evaluación -y, más recientemente, por la contribución de la evaluación a la democracia- no ha hecho sino acrecentarse, como demuestran las últimas obras de House (2006), Stake (2006), Kushner (2005), Kushner y Adelman (2006), MacDonald y Kushner (2005), Schwandt (2003; 2005) o Greene (2001; 2003), por citar sólo algunas. El tema ha logrado hacerse un espacio destacado en los catálogos de estándares y principios-guía de asociaciones o colectivos profesionales -por ejemplo, los de las sociedades suiza y alemana de evaluación⁷ o los del *Joint Committee on Standards for Educational Evaluation* (1994)- o de organizaciones internacionales como el Banco Mundial (2003). Y es que en un contexto como el actual, marcado por la preocupación por el buen gobierno y la calidad de la democracia, la clásica tesis “cualitativista” sólo podía cobrar nueva fuerza.

2. A la búsqueda de otras referencias útiles para construir un esquema de metaevaluación: un recorrido por los principales enfoques metaevaluativos

Como es lógico, el interés por la dimensión política de las evaluaciones se ha dejado sentir también, y de manera notable, en el campo de la metaevaluación. De manera más o menos explícita, la mayoría de las propuestas en este ámbito incorporan criterios sobre la conveniencia de fomentar la participación de los *stakeholders*, hacer de la evaluación un espacio representativo de la pluralidad de intereses en juego, propiciar la negociación y atenerse a unos “mínimos” estándares o principios éticos durante el proceso evaluativo.

A continuación haremos un breve recorrido por algunos de los principales enfoques de evaluación y metaevaluación con el objeto de explorar cuáles son las cuestiones en las que se detienen a la hora de analizar la dimensión política de las evaluaciones.

· La obra de **Michael Scriven (1969; 1991; 2007)** constituye, en este recorrido, una parada ineludible. No en vano, fue este autor el que acuñó el término *metaevaluación* a finales de los años sesenta. En su opinión, el metaevaluador puede y debe servirse no sólo de aquellos estándares que han sido específicamente formulados para la metaevaluación, sino también de aquellos otros que han sido concebidos para la realización de evaluaciones. De esta consideración surgieron sus dos propuestas: la *Meta-Evaluation Checklist* y la *Key Evaluation Checklist*.

La *Meta-Evaluation Checklist* (MEC) incluye una serie de *ítems*, que el autor obtiene de diversas fuentes, sobre el *deber ser* del estudio evaluativo. En la MEC se considera que la evaluación ha de explicitar y justificar los estándares de valor utilizados y ser comprensiva, comprensible, creíble, ética, factible, suficientemente precisa desde el punto de vista técnico -pero sin que ello haga olvidar cuestiones como la relevancia de la información proporcionada-, “sensible” desde el punto de vista político -esto es, que atienda y responda a las necesidades de las distintas *audiencias*-, oportuna, segura, válida y rentable en términos de coste-eficacia.

La *Key Evaluation Checklist* (KEC), en principio elaborada como referencia para la realización de evaluaciones, puede también resultar de utilidad al metaevaluador. La KEC, que el autor ha revisado

y actualizado recientemente (Scriven, 2007), no debe ser considerada como una lista de *desiderata*, sino más bien como un catálogo de cuestiones a las que habría que prestar atención en un proceso metaevaluativo. El catálogo se estructura en cuatro apartados⁸. Bajo la rúbrica “fundamentos”, se recogen las cuestiones relativas a los antecedentes y al contexto de la evaluación, la descripción del programa y la definición de los términos empleados, la identificación de los usuarios (*impactees*), los costes y recursos, y los criterios de valor. El siguiente apartado se dedica a la “subevaluación” de determinados aspectos, como son el propio proceso del programa, los resultados, los costes, la comparación con las alternativas existentes al programa evaluado y la posibilidad de generalización de los resultados de la evaluación. En la última parte, referida a las “conclusiones e impactos”, se realiza una breve síntesis de lo expuesto en apartados anteriores, se recogen las explicaciones sobre los éxitos y fracasos del programa, y las recomendaciones que se entienden pertinentes, la atribución de responsabilidades y la justificación correspondiente, la difusión de la información producida y, finalmente, la metaevaluación.

- **Schwandt y Halpern (1988)** se encuentran también entre los autores que mayor atención han prestado al tema de la metaevaluación. Para estos autores, la metaevaluación, como metáfora de la auditoría fiscal, ha de servir para mejorar la calidad de las evaluaciones. Más concretamente, la metaevaluación debe, en primer término, comprobar la pertinencia de la metodología de la evaluación y, en segundo término, si la información producida es o no correcta. Para ello, ha de ser relevante, fiable, suficiente, representativa y haber sido recabada a tiempo. Sólo si se cumplen estos criterios, las interpretaciones -y los juicios y recomendaciones que se basen en ellas- serán adecuadas.

- Otra de las referencias obligadas en el campo de la metaevaluación es la obra de **Daniel Stufflebeam** (1999). La propuesta más reciente de este autor, la denominada *Program Evaluations Metaevaluation Checklist*, constituye una adaptación al ámbito metaevaluativo de los estándares de evaluación de programas del Joint Committee: utilidad, viabilidad, pertinencia y rigor⁹. Su finalidad es servir de guía para verificar si un proceso evaluativo se atiene a estos estándares, que son considerados el paradigma de la calidad, el *deber ser* de la evaluación. Cada uno de los estándares se divide, a su vez, en una serie de subestándares cuyo cumplimiento se valora a partir de una escala que va del 0-1 -insatisfactorio- al 6 -excelente.

- En una línea argumentativa muy similar se encuentra la denominada *Guiding Principles Checklist for Evaluating Evaluations* de **Stufflebeam, Goodyear, Marquart y Johnson (2005)**, elaborada a partir de los *Guiding Principles for Evaluators* de la *American Evaluation Association* (2004)¹⁰. Precisamente, de los cinco grandes bloques de principios -investigación sistemática, competencia, integridad / honestidad, respeto por las personas y responsabilidad por el bienestar general y el de la sociedad-, tres se refieren a cuestiones de índole ético-política, al igual que en el caso de los *Guiding Principles* en los que se basa.

- “Diálogo” y “transparencia” constituyen los principios clave sobre los que descansa la reflexión de **Uusikylä y Virtanen** (2000). Sostienen estos autores que la bondad y el mérito de las evaluaciones han de ser determinados por todos los *stakeholders*. Ciertas decisiones -como qué tipo de información se requiere o qué criterios han de emplearse para valorar la calidad de una evaluación- no pueden recaer en un único actor. La metaevaluación se concibe, pues, “como un proceso que propicia el diálogo abierto y el juicio colectivo sobre la utilización de los resultados de la evaluación y, de este modo, contribuye al aprendizaje organizativo mediante la investigación evaluativa” (2000: 52).

- Pero, sin duda, es en el *modelo de evaluación democrática y deliberativa* de **House y Howe (2000)** donde la dimensión política de las evaluaciones adquiere mayor protagonismo. Inicialmente concebido para la evaluación, su versatilidad lo ha convertido en una magnífica guía para el metaevaluador, además de ser un auténtico referente para todos aquellos que se interesan por la dimensión ético-política de las evaluaciones. El modelo pretende facilitar la obtención de conclusiones válidas, sin sesgos, allí donde existan opiniones encontradas acerca de un programa. Para ello, las

evaluaciones -o, en su caso, las metaevaluaciones- han de seguir tres principios o criterios básicos: la *inclusión*, el *diálogo* y la *deliberación*. En primer lugar, se entiende que la evaluación ha de *incluir* los intereses, valores y opiniones de los principales *stakeholders* implicados en la política o el programa, aunque no todos tengan idéntica participación en el proceso evaluativo. La evaluación, en segundo término, ha de propiciar el *diálogo* con y entre los *stakeholders*, lo que no implica necesariamente que estos lleguen al final a un acuerdo. El último de los criterios alude al proceso de *deliberación* que ha de desarrollarse en las evaluaciones, con el fin de que puedan extraerse conclusiones válidas y fundamentadas acerca de la política pública que se ha sometido a evaluación (ver resumen en Cuadro 1).

Cuadro 1
Propuestas y “checklists” para la metaevaluación
Cuadro-resumen

	Dimensiones, ítems, estándares
Meta-Evaluation Checklist (Scriven, 1991)	<p>La evaluación debe...</p> <ul style="list-style-type: none"> · explicitar y justificar los estándares de valor utilizados, · ser comprensiva, · ser comprensible, · ser creíble, · ser ética, · ser factible, · ser precisa desde el punto de vista técnico -pero sin que se olviden cuestiones como la relevancia de la información proporcionada-, · ser “sensible” desde el punto de vista político -esto es, que atienda y responda a las necesidades de las distintas <i>audiencias</i>-, · ser oportuna, · ser segura, · ser válida, · ser rentable en términos de coste-eficacia.
Key Evaluation Checklist (Scriven, 1991; 2007)	<p>La metaevaluación debe analizar...</p> <ul style="list-style-type: none"> · los antecedentes y el contexto de la evaluación, · la definición de los términos empleados, · las descripciones de la evaluación que hagan sus responsables, · los criterios de valor, · la identificación de las audiencias y de los <i>stakeholders</i>, · la determinación de los costes y recursos, · la valoración del proceso de investigación y de sus resultados, · las alternativas existentes a la evaluación, · la posibilidad de generalización de la evaluación, · la significación global de la investigación evaluativa, · la existencia de recomendaciones sobre el uso o el destino de la evaluación.
(Schwandt y Halpern, 1988)	<p>La metaevaluación debe comprobar...</p> <ul style="list-style-type: none"> · si la metodología empleada en la metaevaluación es la adecuada, · si las evidencias encontradas son correctas (y, por tanto, relevantes, fiables, suficientes, representativas y han sido recabadas a tiempo).
Program Evaluations Metaevaluation Checklist (Stufflebeam, 1999)	<p>Las evaluaciones deben cumplir los siguientes requisitos:</p> <ul style="list-style-type: none"> · utilidad, · viabilidad, · pertinencia* (adecuación a las normas), · rigor (precisión).
Guiding Principles Checklist for Evaluating Evaluations (Stufflebeam, Goodyear,	<p>La metaevaluación debe analizar...</p> <ul style="list-style-type: none"> · si la evaluación es una investigación sistemática, · el grado de competencia de los evaluadores,

Marquart y Johnson, 2005)	<ul style="list-style-type: none"> · la integridad y honestidad del proceso, · si ha existido respeto por las personas, · si los evaluadores han asumido responsabilidad por el bienestar de la sociedad.
(Uusikilä y Virtanen, 2000)	<p>Los criterios empleados en la metaevaluación se agrupan en...</p> <ul style="list-style-type: none"> · guía (definición del proceso de evaluación y expectativas de la agencia), · selección de los evaluadores, · planificación y desarrollo del proceso evaluativo, · calidad de los informes de evaluación, · contactos y colaboración entre los distintos actores durante el proceso de evaluación, · experiencias de las agencias y los evaluadores sobre el proceso evaluativo, · valoración global por parte de agencias y evaluadores acerca del éxito del proceso evaluativo.
Evaluación democrática deliberativa (House y Howe, 2000)	<p>La metaevaluación debe comprobar si la evaluación se ha atendido a los principios de...</p> <ul style="list-style-type: none"> · inclusión, · deliberación, · diálogo.

* El término en inglés es *propriety*.

Fuente: elaboración propia.

A la vista de todas estas consideraciones, parece evidente que la metaevaluación no puede quedar reducida al mero examen de la calidad técnica de una evaluación. Ciertamente es que las evaluaciones son, en lo esencial, procesos de investigación, pero también lo es que son empresas eminentemente políticas que se desarrollan en un contexto también político. De hecho, como recuerda Patton (1997), la metaevaluación es tanto más necesaria cuanto más politizado esté el contexto y cuanto más “visibilidad” e importancia hayan alcanzado las evaluaciones en él. La actividad metaevaluativa se convierte no sólo en un imperativo profesional para los evaluadores, sino en una empresa de innegable interés para los académicos, los decisores y para la sociedad en su conjunto (Stufflebeam, 2001).

3. Una propuesta de metaevaluación desde la Ciencia Política

El esquema de metaevaluación que proponemos descansa en dos premisas básicas:

a. La metaevaluación ha de reconocer, analizar y enjuiciar tanto la dimensión técnica como la dimensión política de las evaluaciones.

b. Las decisiones que los distintos actores adoptan y las relaciones que se establecen entre ellos afectan e inciden en ambas dimensiones, configurando así los procesos de evaluación.

Nuestra propuesta se construye, por tanto, sobre la base del reconocimiento de la doble dimensión de las evaluaciones. En la denominada *dimensión técnica* cabría incluir las características más evidentes y explícitas de las evaluaciones, aquellas relacionadas con la evaluación en tanto que proceso de investigación: los objetivos formales, el enfoque metodológico, los métodos y las técnicas de recolección y análisis de datos, etc. La *dimensión política* de la evaluación, por su parte, integra aquellos componentes ideológicos y cognitivo-simbólicos asociados a las ideas, valores, opiniones e intereses de los distintos actores así como los vinculados a las relaciones (de poder o de búsqueda del poder) que se establecen entre ellos.

Conviene aclarar que la distinción entre las dos dimensiones es meramente analítica y no “real”. De hecho, muchas de las decisiones adoptadas por los *stakeholders* afectan, en mayor o menor medida, a las dos dimensiones. Así, resulta extremadamente difícil encontrar una cuestión técnica que no tenga implicaciones políticas -por ejemplo, la elección del enfoque metodológico-, de la misma forma que lo es hallar una cuestión considerada *a priori* como política que no tenga consecuencias en el plano técnico -la participación o exclusión de determinados *stakeholders*.

Nuestro esquema, por tanto, no pretende ser un catálogo de estándares que la evaluación debe cumplir, sino, más bien, un conjunto de cuestiones o subdimensiones cuyo análisis permite reconstruir y enjuiciar el proceso evaluativo. Así pues, nos interesan especialmente las decisiones adoptadas por

los actores, las interacciones (políticas) que se establecen entre ellos y cómo esas decisiones e interacciones configuran la evaluación y la medida en que ésta se convierte en un espacio representativo de la pluralidad de visiones existentes sobre un programa -y no sólo de quienes la encargan o financian. El esquema que proponemos pretende aproximarse más a la visión comprensiva y orientada a la experiencia de la que habla Stake (2006) que a los métodos más clásicos y convencionales regidos por criterios y estándares. En nuestra propuesta, la metaevaluación se construye a partir de la experiencia de los actores, del diálogo estrecho, la reflexión y la familiaridad del metaevaluador “con las personas, con los hechos de la evaluación y con las opiniones de la gente” (Stake, 2006: 257).

Además de ello, el esquema permite orientar tanto a los metaevaluadores como a los evaluadores a la hora de identificar qué decisiones pueden contribuir o han contribuido al éxito o al fracaso de un proceso evaluativo. Como es lógico, no se trata, en ningún caso, de buscar a los “culpables” o a los “héroes” de las evaluaciones, sino de detectar aquellas decisiones que afectan o pueden afectar positiva o negativamente la utilidad, la pertinencia, el rigor o la viabilidad de las evaluaciones. El objetivo último es extraer lecciones que permitan una mejor “gestión” de los procesos evaluativos.

Se trata, además, de una propuesta flexible, aplicable tanto en la metaevaluación formativa como en la sumativa, abierta a la inclusión de nuevos ítems y a la utilización de distintas técnicas de recolección y análisis de datos.

La Tabla 1 recoge, de manera gráfica, las subdimensiones que se han de tener en cuenta en la metaevaluación.

Tabla 1
Dimensiones y subdimensiones del esquema de metaevaluación

DIMENSIÓN POLÍTICA	←—————→	▼ DIMENSIÓN TÉCNICA
	Selección del equipo evaluador	
	Existencia y contenido del contrato de evaluación	
	Descripción y contextualización del programa	
	La evaluación ¿tiene en cuenta el contexto en el que se desarrolla?	
	¿Se prevén mecanismos de metaevaluación?	
	¿De qué modo se articulan evaluación y metaevaluación?	
	¿Quiénes participan en el proceso de evaluación?	
	¿De qué modo? ¿Quiénes toman qué decisiones?	
	¿Quiénes quedan excluidos del proceso decisorio?	
	¿Cuáles son las razones que se aducen?	
	¿Se establecen con claridad los roles de cada <i>stakeholder</i> ?	
	Las distintas perspectivas e intereses sobre el programa, ¿están adecuada y suficientemente representadas? ¿Cuáles son los mecanismos empleados para ello?	
	¿Hay desequilibrios de poder entre <i>stakeholders</i> ? ¿De qué modo influye esa circunstancia en la evaluación?	
	¿Se han explicitado los distintos intereses y visiones sobre el programa?	
	Carácter externo o interno de la evaluación	
	¿Cuáles son los objetivos explícitos de la evaluación?	
	¿Cuáles son los objetivos implícitos?	
	¿Qué intereses representan?	

¿Son los objetivos formulados los más adecuados para enjuiciar el programa público?	Oportunidad técnica de la evaluación con respecto al programa
	¿Son adecuadas y suficientes la formación, experiencia y competencia técnica de los evaluadores?
El enfoque metodológico elegido ¿es el más apropiado para analizar y enjuiciar el programa?	
¿Cuáles son los criterios de valor?	Las técnicas de recogida y análisis de datos ¿son las apropiadas teniendo en cuenta los objetivos de la evaluación?
¿Se han hecho explícitos?	La calidad de las técnicas elegidas ¿justifica el coste que entrañan?
	Rigor en la fase de trabajo de campo
	Fiabilidad, rigor y consistencia del análisis de datos
	Las conclusiones ¿se derivan de los datos obtenidos?
	Suficiencia y pertinencia de los recursos empleados
Los clientes ¿admiten comentarios y críticas con respecto al programa?	
¿Es receptivo el equipo evaluador a las sugerencias y comentarios críticos?	
¿Cómo son las relaciones de los distintos <i>stakeholders</i> con el equipo evaluador?	
Oportunidad política de la evaluación (en relación al contexto y al programa evaluado)	
	Grado de compromiso de los distintos <i>stakeholders</i> hacia el proceso de evaluación
¿Cuál es el estilo decisorio predominante?	
¿Negociación y búsqueda del consenso o, por el contrario, estilo impositivo?	
¿Cuál es la actitud del equipo evaluador con respecto a la participación de los <i>stakeholders</i> ?	
	Imparcialidad del equipo evaluador con respecto al programa y a los clientes
	¿Se han arbitrado mecanismos para la gestión, el seguimiento del proceso evaluativo y/o para la resolución de conflictos?
	¿Funcionan satisfactoriamente los circuitos de información y comunicación?
	Los cambios introducidos en el plan inicial de evaluación (en su caso) ¿han sido oportunamente comunicados a los <i>stakeholders</i> ?
	Transparencia con respecto a los costes del proceso de evaluación
Respeto al principio de confidencialidad	
	Claridad y estilo de los informes de evaluación

¿Existen presiones para cambiar el sentido de los datos o el de los informes de evaluación?

Difusión y/o publicación de los resultados de la evaluación

¿Se han ocultado datos importantes a ciertos *stakeholders*?

Los informes de evaluación
¿incluyen información suficiente?

La información producida
¿ha llegado a tiempo de ser utilizada?

¿Proporcionan los informes una descripción
y un juicio equilibrado sobre el programa u ofrecen,
por el contrario, una imagen unitaria, parcial
o distorsionada sobre el mismo?

Utilidad de la evaluación para los distintos *stakeholders*

¿Ha servido la evaluación -tanto si es sumativa como formativa-
para mejorar o conocer con mayor profundidad el programa?

¿Cómo ha influido la evaluación sobre el contexto del programa?

Mejoras introducidas en el programa a raíz de la evaluación

¿Qué medidas se han adoptado para incrementar las posibilidades
de utilización de las evaluaciones y propiciar el aprendizaje?

Impacto de la evaluación en las actitudes y percepciones acerca del programa

¿Cuál es el principal uso al que
se ha destinado la evaluación?

¿Cuáles son los principales “efectos colaterales” de la evaluación?

¿De qué modo han condicionado los “efectos colaterales” el proceso de evaluación?

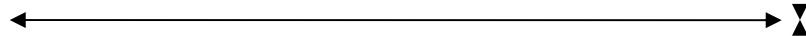
¿Se reconocen y admiten los principales
fallos y limitaciones de la evaluación?

¿Es el proceso de evaluación
representativo y democrático?

Relaciones entre evaluadores y equipo metaevaluador

Impacto de la metaevaluación
en el proceso evaluativo

**DIMENSIÓN
POLÍTICA**



**DIMENSIÓN
TÉCNICA**

4. Un estudio de caso: la metaevaluación del programa de reforma de enseñanzas medias¹¹

El esquema presentado en las páginas precedentes se aplicó en la metaevaluación de las evaluaciones de tres destacados programas de reforma educativa de nivel no universitario. Estos tres programas, como la mayoría de los implantados por los gobiernos socialistas en la España de los ochenta, tuvieron carácter experimental. La finalidad era que la experimentación sirviera de “laboratorio” para la posterior modernización del conjunto del sistema educativo¹².

Tres fueron las evaluaciones objeto del estudio metaevaluativo:

- Evaluación de la reforma de las enseñanzas medias (1984-1992).
- Evaluación del programa de integración de alumnos con necesidades educativas especiales (1985-1989).
- Evaluación del Proyecto Atenea para la introducción de los ordenadores en las aulas (1987-1990).

Las tres evaluaciones fueron íntegramente financiadas con fondos públicos procedentes de la administración central, al igual que los programas a los que se referían. El estudio de metaevaluación tuvo carácter sumativo, esto es, fue realizado una vez que habían concluido las evaluaciones. Hemos de aclarar, igualmente, que la metaevaluación no fue el resultado de un encargo, sino que formó parte de una investigación académica.

No nos detendremos aquí a explicar con detalle el proceso de aplicación del esquema en la metaevaluación realizada. Haremos tan sólo algunos comentarios de carácter general en relación con el proceso de selección de casos y las técnicas de recolección de datos.

Por lo que se refiere a la primera de estas cuestiones, baste señalar que las evaluaciones seleccionadas lo fueron por ser las que tuvieron un mayor predicamento en el ámbito educativo, quizá porque los programas que evaluaban fueron también los de mayor relevancia en la época.

Las técnicas empleadas para la recolección de datos fueron la entrevista semiestructurada y el análisis de documentos. La elección de la entrevista semiestructurada estuvo condicionada no sólo por los problemas que planteaba la utilización de otras técnicas, como los grupos de discusión, sino por su idoneidad para conocer las actitudes, intereses, valores, creencias y expectativas de los actores implicados en los procesos de evaluación, como recuerdan acertadamente autores como Fielding (1994) o Alonso (1994). El análisis de documentos -principalmente de los informes de evaluación aunque también de documentos oficiales del Ministerio de Educación y Ciencia, de la legislación educativa o de la bibliografía especializada- nos permitió obtener importantes referencias sobre las evaluaciones analizadas al tiempo que ampliar y contrastar los datos obtenidos en las entrevistas.

Así pues, a través de estas dos técnicas recabamos información acerca de todas y cada una de las subdimensiones contempladas en nuestro esquema de metaevaluación. Con objeto de llegar a unas conclusiones suficientemente fiables recurrimos a la triangulación¹³ tanto de los datos obtenidos de los distintos entrevistados como de los procedentes de las dos técnicas empleadas. Entendemos que la triangulación es particularmente útil no sólo en la evaluación sino también en la metaevaluación. Suscribimos, pues, la opinión de Stake (2006: 262) cuando afirma que con este método “siempre se gana, porque si produce una confirmación, se podrá depositar una mayor confianza en la interpretación y, si por el contrario, produce un desmentido, puede hacer que aumente el reconocimiento de la existencia de diversidades y realidades múltiples”.

Dada la complejidad de la metaevaluación realizada, para los propósitos de este artículo hemos seleccionado sólo una de las evaluaciones que se analizaron, la correspondiente al programa de reforma experimental de enseñanzas medias. Esta evaluación puede considerarse como el primer gran estudio evaluativo sobre programas educativos llevado a cabo en España. Al mismo tiempo y, dadas las limitaciones de espacio, daremos cuenta, únicamente, de los aspectos más destacables del proceso de evaluación y de las principales conclusiones que se derivaron del mismo. Por esta misma razón, omitiremos los nombres de los principales *stakeholders*, así como las fechas en las que fueron realizadas las entrevistas.

4.1 Una breve descripción del programa evaluado

Las enseñanzas medias constituían, en la España de principios de los años ochenta, una de las áreas del sistema educativo más necesitadas de reforma. Los cambios políticos, sociales y económicos que se habían ido produciendo en la sociedad española en la década de los setenta pusieron de relieve las deficiencias de la Ley General de Educación¹⁴ (en adelante, LGE), vigente desde 1970: su engarce con la enseñanza primaria (la denominada EGB, Educación General Básica) y con la enseñanza universitaria, la existencia de enfoques pedagógicos excesivamente tradicionales o la deficiente calidad de la formación profesional.

La reforma de las enseñanzas medias se desarrolló entre 1983 y 1992 y tuvo, como ya hemos mencionado, carácter experimental. Ello permitía a la administración educativa poder contrastar la

teoría y la práctica, y disponer de datos fiables acerca de la viabilidad de sus propuestas antes de su definitiva generalización. Además, la voluntad del equipo ministerial era que el proceso contara con la participación de todos los sectores de la comunidad educativa. Por esta razón, se ofreció a los centros la posibilidad de desarrollar y concretar las propuestas establecidas por el Ministerio mediante la implantación de sus propios proyectos de reforma.

4.2 El proceso de evaluación

La evaluación externa de la reforma experimental de enseñanzas medias estuvo integrada por tres estudios parciales¹⁵ (Álvaro Page ...[et al], 1988; 1990; 1992). No fue un estudio encargado por el responsable del programa, sino el fruto de la iniciativa personal de un técnico del CIDE -un centro de investigación y documentación adscrito al Ministerio de Educación y Ciencia. Este Centro asumió la realización de los trabajos de evaluación, bajo la dirección del técnico, así como su financiación.

Aunque el director de la evaluación no desempeñaba ningún cargo de confianza política, sintonizaba “fuertemente” con los principios y el planteamiento de la reforma, y contaba con el respaldo de su superior, quien, a su vez, se encontraba muy próximo al entonces Ministro de Educación y Ciencia. Estas circunstancias alejaban la posibilidad de una utilización de los datos que fuera perjudicial para los altos cargos ministeriales. La evaluación constituía, además, una propuesta claramente innovadora para la época y una oportunidad para la legitimación del programa ante quienes participaban en él y ante el resto de la comunidad educativa.

La evaluación de la reforma de enseñanzas medias fue una investigación de una enorme complejidad tanto por el elevado número de centros y de alumnos participantes en el programa como por la ingente cantidad de datos producidos y por su propia duración temporal -1984 a 1992. Es lógico, por tanto, que el coste fuera también ciertamente considerable.

La apuesta por un enfoque excluyente de evaluación por objetivos. El objetivo formal y explícito de la evaluación de la reforma experimental de enseñanzas medias, tal y como fue definido por el director del estudio y por su superior jerárquico, era comparar los resultados obtenidos, en determinadas variables, por los alumnos que seguían el plan experimental y por los que cursaban el plan establecido en la Ley General de Educación. La evaluación de la reforma experimental de enseñanzas medias pretendía, en definitiva, evaluar los resultados alcanzados por los alumnos en el cumplimiento de los objetivos académicos marcados por la administración educativa. La evaluación era, en realidad, un estudio de *assessment*.

En línea con su formación académica y profesional y con la perspectiva cuantitativista por la que se decantaban, el director de la evaluación y el director del Centro optaron por un enfoque clásico de evaluación por objetivos y por un diseño de corte cuasi-experimental. En estas decisiones, como en el resto de las que se adoptaron en el transcurso de la evaluación, no tomaron parte ni el responsable del programa ni ningún otro miembro del mismo. De esa forma, ambos actores se convirtieron, desde un primer momento, en los actores centrales del proceso evaluativo.

Sin duda, una evaluación referida al diseño y la implantación del programa hubiera proporcionado una información más rica y exhaustiva -y, por ello, más comprometida- que la que se obtenía de una evaluación que atendía única y exclusivamente a los resultados de los alumnos. El director del CIDE era plenamente consciente de ello: “Creíamos que eso (una evaluación del diseño y la implantación) lo tenían que hacer ellos, la parte interna. Primero porque (nosotros) no teníamos medios (para hacerlo) y si hubiéramos tenido los medios no lo hubiéramos hecho porque hubiera sido visto como una injerencia intolerable en la marcha de la reforma”. La evaluación que se puso en marcha no era quizá la más adecuada desde el punto de vista científico pero, sin duda, era la menos arriesgada desde el punto de vista político. De hecho, a nuestro juicio, una de las principales debilidades de este proceso de evaluación no fue tanto su excesivo cuantitativismo cuanto la identificación del programa con el rendimiento de los alumnos.

Una de las consecuencias derivadas de la elección de un enfoque de evaluación por objetivos fue la exclusión de ciertos actores -los centros educativos o el responsable del programa- de los procesos decisorios de la evaluación. El director de la evaluación y el responsable del Centro no contemplaron en ningún momento la posibilidad de requerir su participación como decisores, ni siquiera como informantes. No parece, sin embargo, que con esta exclusión se pretendiera evitar la presencia de perspectivas potencialmente críticas, sino que respondió únicamente a las preferencias metodológicas de ambos actores.

Como es lógico, la percepción de los distintos *stakeholders* acerca del enfoque y el diseño de la evaluación difería notablemente según su posición y sus intereses. Así, mientras que algunos de los miembros del equipo de evaluación los valoraron muy positivamente, otros evaluadores o los profesores participantes en la reforma fueron muy críticos en este punto. Para ellos, el enfoque elegido llevó a la obtención de unas estadísticas “excesivamente generales”, que “profundizaban poco”, “se quedaban en el puro dato” y “en las que resultaba difícil reconocerse”. La gran mayoría señaló que hubiera preferido un método más cualitativo y participativo que le hubiese permitido conocer el programa con mayor detalle.

La evaluación fue también objeto de críticas por parte de algunos expertos ajenos al proceso. Alvira (1991: 53), por ejemplo, considera que este estudio constituye “el típico ejemplo en el que una evaluación de la implementación era necesaria”. En opinión de Angulo Rasco (1993: 9), la evaluación de la reforma de enseñanzas medias fue una buena muestra de las ideas imperantes en nuestra administración, según las cuales “evaluar suele limitarse a evaluar los puntos terminales (del sistema): centros de enseñanza, docentes y alumnos”.

La aparente ausencia de criterios de valor. En la evaluación de la reforma experimental de enseñanzas medias no se formularon criterios de valor. No obstante, lo cierto es que, en la práctica, el éxito de la reforma se medía a través de los resultados de los alumnos. Los evaluadores manejaban, aunque no fueran conscientes de ello y no lo hicieran explícito, un criterio de valor fundamental. La reforma era tanto mejor cuanto mejores fueran los resultados de los alumnos. De esta forma, se inducía a que los juicios derivados de las conclusiones de la evaluación se basaran tan sólo en los logros del alumnado.

El supuesto carácter externo de la evaluación. La evaluación de la reforma de enseñanzas medias fue formalmente definida como externa. Los responsables del Centro del que formaba parte el equipo evaluador eran conscientes de los reparos que esta definición podía despertar, como quedó patente en el prólogo del segundo de los informes (Álvaro Page ...[et al], 1990: 17). En nuestra opinión, el pretendido carácter externo de la evaluación es ciertamente discutible. En realidad no existía una vinculación directa entre los evaluadores y la unidad responsable del programa. Pero no lo es menos que ambos trabajaban conforme a unas mismas directrices políticas y que el director de la evaluación era un claro partidario de la reforma.

El procedimiento de muestreo y las técnicas de recolección de datos. En cada uno de los tres estudios de la evaluación de la reforma experimental de enseñanzas medias se constituyeron tres grupos: uno experimental y dos de control (uno interno y otro externo). Los grupos experimentales estaban formados, como es obvio, por los alumnos que seguían el plan experimental. Los grupos de control interno se formaron con aquellos alumnos de centros experimentales que cursaban el plan antiguo, mientras que los de control externo se constituyeron a partir de una muestra representativa, a nivel estatal, de sujetos que seguían el plan de la LGE en centros no experimentales. El diseño de la evaluación tuvo, además, carácter *longitudinal*.

Debido al elevado número de centros y alumnos participantes en el programa, se decidió que la aplicación de las pruebas se realizase, en algunos casos, sobre una muestra de cada uno de los tres grupos. De nuevo, fue el director de la evaluación el que tomó esta decisión y quien estableció los criterios que habrían de seguirse para el muestreo, cuyo diseño corrió a cargo de unos expertos

contratados al efecto. El procedimiento de muestreo se llevó a cabo en dos fases. En una primera se procedía a la selección de los centros, mientras que, en la segunda, se seleccionaban los alumnos que integrarían las muestras.

Ciertamente, y a juicio de los expertos entrevistados, el procedimiento de selección de las muestras se ajustó a las características de las poblaciones sobre las cuales se efectuaba el estudio. Sin embargo, el número de centros y alumnos considerados fue extraordinariamente elevado. Ello repercutió de forma negativa en el desarrollo de la evaluación, al incrementar de modo considerable la cantidad de información que había de recabarse y analizarse y, consiguientemente, el coste de la misma.

En coherencia con la perspectiva cuantitativa que dominaba el estudio, los tests y los cuestionarios fueron las principales técnicas empleadas en la recogida de datos. La medición de la mayoría de las variables no planteaba grandes dificultades puesto que existían ya en el mercado pruebas suficientemente contrastadas. En el caso de las variables para las que no se disponía de un test apropiado fue necesario diseñar la prueba *ex novo*, tarea para la que se recurrió a profesores especialistas en las distintas materias.

El desarrollo del trabajo de campo y el tratamiento de la información. La dimensión y la envergadura de la evaluación experimental de la reforma de enseñanzas medias hicieron que el Centro se viera obligado a recurrir a empresas especializadas a la hora de desarrollar el trabajo de campo, ya que no disponía ni del personal ni de los medios materiales necesarios para ello. El equipo evaluador se ocupó de fijar las condiciones en las que habían de realizarse las pruebas.

El tratamiento estadístico de los datos obtenidos se centró en el control de las variables contaminantes. Por iniciativa del director del Centro, la explotación y el análisis de los datos obtenidos tras la aplicación de las pruebas correspondió a diferentes especialistas, que fueron contratados con este fin previa convocatoria de concurso público. Se pretendía evitar que la interpretación de los datos se basara en una única perspectiva y paliar, en lo posible, el hecho de que sólo se hubieran utilizado técnicas cuantitativas.

Los desequilibrios de poder y los actores excluidos. Los centros fueron, junto con el director del programa, los grandes ausentes del proceso de evaluación. Además, la recolección de datos les ocasionó, como señalaron los profesores entrevistados, un sinnúmero de problemas. La realización de las pruebas distraía la atención de los alumnos y restaba tiempo a las actividades habituales del aula; estas circunstancias, junto con el marcado carácter cuantitativo del estudio, hicieron que su imagen de la evaluación fuera tan negativa.

Ni siquiera la entrada de otros actores en el escenario de la evaluación -los responsables del muestreo, los profesores que diseñaron los tests para la recolección de datos, las empresas encargadas del trabajo de campo o los expertos contratados para la explotación de los datos- consiguió alterar en modo alguno la distribución de poder que se había establecido en los inicios del proceso evaluativo. El director del equipo evaluador y el director del Centro siguieron teniendo en sus manos el control de las principales decisiones. La participación de los actores excluidos hubiera contribuido, sin duda, a que la evaluación ofreciera una imagen más completa y fidedigna del programa de la que mostró.

Las principales conclusiones y el estilo de los informes de evaluación. Para el equipo de evaluación, los resultados que arrojó la comparación entre alumnos de la reforma experimental y alumnos que cursaban el plan establecido por el plan antiguo eran claramente favorables a la reforma. En los tres informes de evaluación se afirmaba que, aunque las diferencias entre los resultados obtenidos por los dos tipos de alumnos no eran muy grandes, sí podían calificarse de significativas. Sin embargo, de su lectura y de los comentarios de algunos de los expertos entrevistados, se deduce que esos resultados no eran tan concluyentes como se afirmaba. Muy probablemente, ello esté relacionado con las actitudes que quienes diseñaron la evaluación tenían con respecto al programa experimental.

El enfoque del estudio se reflejó claramente en los informes de evaluación. Nos encontramos con unos informes prototípicos de las evaluaciones de corte cuantitativista, de estilo eminentemente técnico y cuya lectura y comprensión requieren de ciertos conocimientos estadísticos. No se trata de informes accesibles, pensados y dirigidos al gran público o a la comunidad educativa, que cumplan con el criterio de “claridad” establecido por el *Joint Committee*, lo que no deja de ser una prueba más del carácter que se pretendió conferir a la evaluación.

La utilidad y el impacto de la evaluación. La valoración que el responsable del programa y los profesores participantes en el mismo hicieron acerca de la utilidad de la evaluación de la reforma de enseñanzas medias fue claramente negativa.

Para el director del programa, uno de los grandes ausentes del proceso, la evaluación externa fue una experiencia carente de interés y de utilidad. Si recibió su visto bueno fue porque no interfería en el normal desarrollo del programa y porque los datos que iba a producir no podían constituir amenaza alguna. Sin duda, su participación en el proceso hubiera permitido obtener información de mayor utilidad para la mejora del proceso experimental.

El escaso interés que inicialmente pudo albergar el profesorado por la evaluación fue disminuyendo a medida que comprobaba la poca utilidad de sus resultados. Pese a esta circunstancia, los docentes colaboraron sin mayores problemas en la investigación. Seguramente, su actitud hubiese sido mucho más crítica si hubiesen sido “objeto de evaluación”, y si ésta hubiera entrado en conflicto con sus intereses profesionales.

Por el contrario, el impacto de la evaluación sobre el Centro en el que se desarrolló fue muy positivo. La realización de la evaluación externa le proporcionó no sólo recursos económicos sino también prestigio y una inmejorable oportunidad para su consolidación como unidad especialista en el ámbito de la evaluación educativa, dentro y fuera del Ministerio¹⁶.

A modo de reflexión final

La visión de la metaevaluación que defendemos como análisis de actores lleva implícita, a su vez, una determinada imagen de lo que es un proceso de evaluación. A nuestro juicio, los procesos evaluativos han de ser considerados, y por tanto analizados, como una sucesión de decisiones vinculadas a las dimensiones técnica y política, a través de las cuales los actores van “construyendo” dichos procesos y confiriéndoles sus características distintivas.

Toda propuesta teórica requiere de una contrastación empírica. El esquema de metaevaluación expuesto en este artículo permite:

- Recuperar, al igual que el enfoque de políticas públicas, el interés por los actores (lo que no supone apostar por ellos en detrimento de la estructura)¹⁷.
- Evidenciar cómo las decisiones de los actores -determinadas, a su vez, por sus ideas, valores, preferencias e intereses- condicionan y modulan los procesos de evaluación. Del mismo modo, nos ofrece las claves para comprender por qué fracasan algunos procesos evaluativos -tremendamente costosos e “impecables” desde el punto de vista técnico- o por qué sus resultados no son utilizados.
- Descubrir los desequilibrios de poder en el seno de la evaluación, qué *stakeholders* controlan la adopción de las principales decisiones, y cómo todo ello afecta al desarrollo y los resultados de las evaluaciones.
- Comprobar cómo la falta de participación de ciertos actores hace que la evaluación presente una imagen parcial e incompleta del programa, e influye negativamente en la utilidad y en la utilización del proceso evaluativo.
- Poner de manifiesto, en definitiva, el interés que la evaluación de políticas públicas puede tener para una disciplina como la Ciencia Política.

La aplicación del esquema de análisis aquí presentado ha resultado ser de gran utilidad a la hora de profundizar en el análisis de la dimensión política de las evaluaciones. Creemos que, con nuevas

aportaciones y con la necesaria adaptación al contexto de las políticas evaluadas, puede serlo también en otros de similares características.

Notas

¹ Nuestra reflexión se refiere exclusivamente a aquellas evaluaciones de programas y políticas públicas que se financian con fondos públicos.

² Ello no nos debe llevar, como nos previene Morgan (1993: 212), “a ver política por todas partes y a buscar ‘agendas escondidas’ donde no hay agendas”. Así pues, el hecho de que las evaluaciones tiendan o puedan ser instrumentalizadas políticamente, no debe llevar al estudioso a partir de una aproximación *patológica* en su análisis. El que una evaluación no sea utilizada tal y como fue previsto, o el que su diseño técnico no resulte ser el más apropiado para los objetivos propuestos, no quiere decir necesariamente que detrás de todo ello haya una intencionalidad política ilegítima.

³ Con objeto de no alargar excesivamente la exposición, no incluiremos aquí la relación de los trabajos más destacables en cada uno de estos temas. Remitimos, para ello, al lector a la obra ya citada (García Sánchez, 2005).

⁴ La definición de “actor” de la que partimos se encuentra en García Sánchez (2007).

⁵ Recordemos que, desde David Easton (1965), la política se concibe como aquella actividad orientada a la regulación de los conflictos sociales mediante decisiones que tienen carácter vinculante para el conjunto de la sociedad. A través de esas decisiones se otorga valor a determinados demandas o intereses en detrimento de otros. Así, un acto se convierte en político cuando se asocia o tiene que ver con dichas decisiones, con las asignaciones imperativas de valor en una sociedad o en un contexto dado.

⁶ Como apunta esta autora, entre los partidarios de esta tesis hay dos posturas muy claras. Por un lado, la de quienes consideran que el evaluador debe comprometerse explícitamente con alguna de las opciones ideológicas que afloran en el proceso de evaluación, discriminar positivamente a aquellos *stakeholders* que se encuentren en posición de desventaja e intentar influir sobre el contexto político induciendo cambios en el mismo (el enfoque de *empowerment* o la evaluación democrática). Por otro, aquellos que apuestan por el pluralismo, por integrar las distintas perspectivas en la evaluación pero *sin tomar partido* por ninguna de ellas, con objeto de evitar sesgos que puedan poner en entredicho la credibilidad del proceso (el modelo de evaluación “sensible” o *responsive evaluation*, o los centrados en la utilización).

⁷ Disponibles, respectivamente, en las direcciones de Internet http://www.seval.ch/en/documents/SEVAL_Standards_2000_en.pdf y <http://www.degeval.de/calimero/tools/proxy.php?id=71> (ambas consultadas el 20 de octubre de 2008).

⁸ En el primero de ellos, titulado “cuestiones preliminares”, se incluyen el informe ejecutivo, la introducción y la metodología de la evaluación. En estricto sentido, sería más un epígrafe del informe de evaluación que un bloque de cuestiones que deban abordarse durante el proceso evaluativo.

⁹ En el proceso de revisión de estos estándares, que se está desarrollando actualmente, se ha añadido un quinto estándar, “metaevaluación” que se desagrega, a su vez, en tres subestándares: M1 “las metaevaluaciones deben ser ‘sensibles’ a las necesidades de los usuarios potenciales”; M2 “las metaevaluaciones deben identificar y aplicar estándares de calidad pertinentes”; y M3 “las metaevaluaciones deben basarse en información precisa y adecuada”. Puede encontrarse información adicional sobre este proceso de revisión en la página web del *Joint Committee*, <http://www.wmich.edu/evalctr/jc/> (consultada el 12 de octubre de 2008).

¹⁰ La primera versión de estos *Guiding Principles* data de 1994, cuando la Junta de la *American Evaluation Association* (AEA) constituyó un grupo de trabajo con el propósito de elaborar un conjunto de principios orientadores para la evaluación. En 2002-2003 se constituyó un nuevo equipo de trabajo

que revisó la primera versión y cuyo informe, una vez aprobado por los miembros de la AEA, dio lugar a los *Guiding Principles* de 2004.

¹¹ En García Sánchez (2005) puede encontrarse toda la información relativa a la aplicación del esquema de metaevaluación a los estudios de caso.

¹² De hecho, la estructura del sistema prevista en la Ley de Ordenación General del Sistema Educativo (LOGSE) y vigente desde 1990 hasta 2002 fue, en gran medida, deudora de aquellos primeros programas experimentales.

¹³ Recordemos que la estrategia de validación convergente, también conocida como “triangulación”, tal y como fue definida por Campbell (1956), hace referencia al empleo de distintas técnicas de recogida de datos para medir un mismo concepto. Denzin (1978) amplió el significado del término para significar el uso no sólo de distintas técnicas de recolección de datos, sino también de distintas teorías, distintos investigadores, distintas metodologías (o bien la combinación de todos estos elementos) con el fin de evitar o paliar las diferencias que presenta cada uno de ellos y reforzar la fiabilidad de la información producida. En el ámbito de la evaluación, Stake (2006) va aun más allá al considerar que los datos pueden también triangularse con respecto al tiempo, el espacio y las personas.

¹⁴ Ley 14/1979, de 4 de agosto, General de Educación y Financiamiento del Sistema Educativo.

¹⁵ Las diferencias entre ellos no eran sustanciales por lo que, a efectos de la metaevaluación, fueron considerados como una única evaluación.

¹⁶ De hecho, en el transcurso de la evaluación se creó, dentro del Centro, el Servicio de Evaluación. El director de la evaluación de la reforma de enseñanzas medias fue nombrado responsable del mismo.

¹⁷ Véase, a propósito del debate *agencia-estructura*, García Sánchez (2007).

Bibliografía

- Alonso, L. E. (1994), “Sujeto y discurso: el lugar de la entrevista abierta en las prácticas de la sociología cualitativa”, en *Métodos y técnicas cualitativas de investigación en ciencias sociales*, J. M. Delgado y J. Gutiérrez (coords.), Madrid, Síntesis.
- Álvaro Page, M. ...[et al] (1988), *Evaluación externa de la reforma experimental de enseñanzas medias (I)*, Madrid, Centro de Investigación y Documentación Educativa.
- _____ (1990), *Evaluación externa de la reforma experimental de enseñanzas medias (II)*, Madrid, Centro de Investigación, Documentación y Evaluación.
- _____ (1992), *Evaluación externa de la reforma experimental de enseñanzas medias (III)*, Madrid, Centro de Investigación, Documentación y Evaluación.
- Alvira, F. (1991), *Metodología de la evaluación de programas*, Madrid, Centro de Investigaciones Sociológicas.
- Angulo Rasco, J. F. (1993), “Evaluación del sistema educativo: algunas respuestas críticas”, en *Cuadernos de Pedagogía*, N° 219, Madrid, pp. 8-15.
- American Evaluation Association (2004), *Guiding Principles for Evaluators*, Fairhaven, AEA, <http://www.eval.org/Publications/GuidingPrinciples.asp>, 10-10-2008.
- Banco Mundial (2003), *Independent Evaluation: Principles, Guidelines and Good Practice*, Washington, Banco Mundial (Development Grant Facility Technical Note), mimeo.
- Campbell, D. T. (1956), *Leadership and its Effects upon the Group*, Columbus, Ohio State University Press.
- Delgado, J. M. y Gutiérrez, J. (coords.) (1994), *Métodos y técnicas cualitativas de investigación en ciencias sociales*, Madrid, Síntesis.
- Denzin, N. K. (1978), *The Research Act*, New York, McGraw-Hill.
- Easton, D. (1965), *A Framework for Political Analysis*, Englewood Cliffs, Prentice-Hall.

- Fielding, N. (1994), "Qualitative Interviewing", en *Researching Social Life*, N. Gilbert (ed.), London, Sage.
- García Sánchez, E. (2005), *La evaluación de programas de reforma educativa en España: tres estudios de caso desde un enfoque de metaevaluación*, Madrid, Universidad Complutense de Madrid. Tesis doctoral.
- _____ (2007), "El concepto de actor: reflexiones y propuestas para la ciencia política", en *Andamios*, Vol. 3 N° 6, México, pp. 199-216.
- Gilbert, N. (ed.) (1994), *Researching Social Life*, London, Sage.
- Greene, J. (2001), "Dialogue in Evaluation: a Relational Perspective", en *Evaluation*, Vol. 7 N° 2, London, pp. 181-187.
- _____ (2003), "War and Peace... and Evaluation", en *Studies in Educational Policy and Educational Philosophy*, N° 2, Uppsala, <http://www.upi.artisan.se/docs/Doc186.pdf>, 12-10-2008.
- Guba, E. G. y Lincoln, Y. S. (1981), *Effective Evaluations: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*, San Francisco, Jossey-Bass.
- _____ (1987), "The Countenances of Fourth-Generation Evaluation: Description, Judgement and Negotiation", en *The Politics of Program Evaluation*, D. J. Palumbo (ed.), Beverly Hills, Sage.
- _____ (1989), *Fourth Generation Evaluation*, Newbury Park, Sage.
- Hamilton, D. (ed.) (1977), *Beyond the Numbers Game*, London, MacMillan.
- Harvard Family Research Project (2005), "A Conversation with Jennifer Greene", en *The Evaluation Exchange*, Vol. 11 N° 3, Cambridge, Fall, pp. 10-11.
- House, E. R. (1980), *Evaluating with Validity*, Beverly Hills, Sage.
- _____ (1993), *Professional Evaluation: Social Impact and Political Consequences*, Newbury Park, Sage.
- _____ (2006), "Democracy and Evaluation", en *Evaluation*, Vol. 12 N° 1, London, pp. 119-127.
- House, E. R. y Howe, K. R. (1999), *Values in Evaluation and Social Research*, Thousand Oaks, Sage.
- _____ (2000), Deliberative Democratic Evaluation Checklist, Kalamazoo, Western Michigan University. The Evaluation Center, http://www.wmich.edu/evalctr/checklists/dd_checklist.PDF, 14-09-2008.
- Joint Committee on Standards for Educational Evaluation (1994), *The Program Evaluation Standards*, Thousand Oaks, Sage.
- Knorr-Cetina, K. D. y Cicourel, A. V. (eds.) (1981), *Advances in Social Theory and Methodology: Towards an Integration of Micro- and Macro-Sociologies*, London Routledge.
- Kushner, S. (2005), "Democratic Theorizing: from Noun to Participle", en *American Journal of Evaluation*, Vol. 26, Los Angeles, pp. 579-581.
- Kushner, S. y Adelman, C. (2006), "Program Evaluation: a Democratic Practice", en *Handbook of Complementary Methods in Education Research*, J. L. Green, G. Camilli y P. B. Elmore (eds.), London, Routledge.
- Leavitt, H. ...[et al] (1974), *Organizations of the Future: Interaction with the Environment*, New York, Praeger Publishers.
- MacDonald, B. (1977), "A Political Classification of Evaluation Studies", en *Beyond the Numbers Game*, D. Hamilton (ed.), London, MacMillan.
- MacDonald, B. y Kushner, S. (2005), "Democratic Evaluation", en *Encyclopedia of Evaluation*, S. Mathison (ed.), Thousand Oaks, Sage.
- Marsh, D. y Stoker, G. (eds.) (1997), *Teoría y métodos de la ciencia política*, Madrid, Alianza Editorial.
- Mathison, S. (ed.) (2005), *Encyclopedia of Evaluation*, Thousand Oaks, Sage.

- Morgan, G. (1993), "Intereses, conflictos y poder: las organizaciones como sistemas políticos", en *Lecturas de teoría de la organización*, C. Ramió y X. Ballart (eds.), Madrid, Ministerio para las Administraciones Públicas, Boletín Oficial del Estado. v.2.
- Palumbo, D. J. (ed.) (1987), *The Politics of Program Evaluation*, Beverly Hills, Sage.
- Patton, M. Q. (1997), *Utilization-Focused Evaluation*, Thousand Oaks, Sage.
- Ramió, C. y Ballart, X. (eds.) (1993), *Lecturas de teoría de la organización*, Madrid, Ministerio para las Administraciones Públicas, Boletín Oficial del Estado. v.2.
- Schwandt, T. A. (2003), "In Search of the Political Morality of Evaluation Practice", en *Studies in Educational Policy and Educational Philosophy*, N° 2, Uppsala, <http://www.upi.artisan.se/docs/Doc183.pdf>, 15-10-2008.
- _____ (2005), "Politics of Evaluation", en *Encyclopedia of Evaluation*, S. Mathison (ed.), Thousand Oaks, Sage.
- Schwandt, T. A. y Halpern, E. S. (1988), *Linking Auditing and Metaevaluation: Enhancing Quality in Applied Research*, Newbury Park, Sage.
- Scriven, M. (1969), "An Introduction to Meta-Evaluation", en *Educational Product Report*, N° 2, pp. 36-38.
- _____ (1991), *Evaluation Thesaurus*, Newbury Park, Sage.
- _____ (2007), Key Evaluation Checklist, Kalamazoo, Western Michigan University. The Evaluation Center, http://www.wmich.edu/evalctr/checklists/kec_feb07.pdf, 20-09-2008.
- Stake, R. E. (2006), *Evaluación comprensiva y evaluación basada en estándares*, Barcelona, Editorial Graó.
- Stufflebeam, D. L. (1999), Program Evaluations Metaevaluation Checklist (Based on the Program Evaluation Standards), Kalamazoo, Western Michigan University. The Evaluation Center, http://www.wmich.edu/evalctr/checklists/program_metaeval_10point.pdf, 21-09-2008.
- _____ (2001), "The Metaevaluation Imperative", en *American Journal of Evaluation*, Vol. 22 N° 2, Los Angeles, pp. 183-209.
- Stufflebeam, D. L.; Goodyear, L.; Marquart, J.; y Johnson, E. (2005), Guiding Principles Checklist for Evaluating Evaluations: Based on the 2004 Guiding Principles for Evaluators, Kalamazoo, Western Michigan University. The Evaluation Center, <http://www.wmich.edu/evalctr/checklists/guidingprinciples2005.pdf>, 21-09-2008.
- Uusikylä, P. y Virtanen, P. (2000), "Meta-Evaluation as a Tool for Learning: a Case Study of the European Structural Fund Evaluations in Finland", en *Evaluation*, Vol. 6 N° 1, London, pp. 50-65.
- Vedung, E. (1997), *Evaluación de políticas públicas y programas*, Madrid, Ministerio de Trabajo y Asuntos Sociales.
- Wardell, M. L. y Turner, S. P. (eds.) (1986), *Sociological Theory in Transition*, London, Allen and Unwin.
- Weiss, C. H. (1987), "Where Politics and Evaluation Research Meet", en *The Politics of Program Evaluation*, D. J. Palumbo (ed.), Beverly Hills, Sage.