

Metaevaluación: entre la calidad técnica y la calidad democrática de las evaluaciones

1. Definiendo el concepto de metaevaluación

Por analogía con el término evaluación¹, el concepto de *metaevaluación* se asocia a aquella investigación sistemática cuyo objetivo es emitir un juicio acerca de la bondad, el valor, los méritos (o deméritos) de una evaluación².

Desde la primera formulación realizada por Scriven, en 1969, este concepto ha ido perfilándose gracias a las aportaciones de éste y otros autores, como Lessinger (1970), Provus (1971), Seligman (1973) o Stufflebeam (1981, 1993). A pesar de que, en los últimos años, ha conocido un auge considerable, su desarrollo no es comparable al de la evaluación. Prueba de ello, es que algunas de las metaevaluaciones más significativas que se han llevado a cabo hasta la fecha,

no son sino aplicaciones de modelos teóricos elaborados para la realización de evaluaciones³.

El propósito de quienes han reflexionado acerca del concepto de *metaevaluación* es, ante todo, encontrar una herramienta que permita valorar el proceso de evaluación y verificar el grado de validez de sus resultados y conclusiones. La evaluación se convierte, de esta forma, en el objeto de estudio del metaevaluador.

2. Un paso más en la definición del concepto *metaevaluación*: las aportaciones de cuantitativistas y cualitativistas

No obstante, aquella definición plantea problemas de aplicación práctica: ¿qué se entiende por "bondad, valor o méritos" de una evaluación? ¿cuáles son los criterios conforme a los cuales ha de juzgarse una evaluación? Las respuestas que las diferentes corrientes metodológicas vienen ofreciendo a esta cuestión son bien distintas.

En líneas generales, las corrientes *cuantitativistas* (también llamadas científistas o positivistas) tienden a hacer hincapié en los aspectos técnico-metodológicos y, en concreto, a hacer de las técnicas cuantitativas la clave esencial de la evaluación⁴.

A juicio de los representantes de estas corrientes, el valor de una evaluación sólo queda plenamente garantizado con el empleo de procedimientos cuantitativos para la recogida y el tratamiento de datos en el marco de un diseño de naturaleza experimental⁵. La metaevaluación tiene, pues, como objetivo calibrar en qué medida el estudio de evaluación se aproxima a este "modelo".

El trabajo de metaevaluación de Bernstein y Freeman (1975) constituye uno de los mejores ejemplos de lo que los enfoques científistas consideran como "calidad" en la investigación evaluativa. Estos autores desagregaron el concepto de calidad en una serie de dimensiones -tipo de muestreo, análisis de datos, procedimientos estadísticos y diseño de la investigación-

¹ Para Rossi y Freeman, la investigación evaluativa consiste en "la aplicación sistemática de los procedimientos de la investigación social, para valorizar la conceptualización y el diseño, la ejecución y la utilidad de los programas de intervención social" (1989: 14). Patton (1986) considera la evaluación de programas como el proceso sistemático de recogida de información acerca de las actividades, las características y los resultados de un programa, de cara a su utilización para reducir incertidumbres, mejorar la efectividad y tomar decisiones sobre el programa. En opinión de Stufflebeam y Shinkfield (1993) y, en la línea de la definición adoptada por el *Joint Committee on Standards for Educational Evaluation*, la evaluación consiste en el enjuiciamiento sistemático de la valía o el mérito de un objeto. Para Carol Weiss la evaluación es "la valoración sistemática del desarrollo y/o los resultados de un programa o una política, en función de un conjunto implícito o explícito de estándares, como un medio de contribuir a la mejora del programa o la política" (1998: 4). Como puede apreciarse, todas las definiciones de *evaluación* coinciden en destacar tres elementos: su carácter de investigación, su finalidad valorativa y su orientación aplicada. Sobre las diferencias entre investigación y evaluación véase Glass y Worthen (1971).

² Este artículo se centra exclusivamente en la evaluación y metaevaluación de programas y proyectos, aunque algunas de las reflexiones que se realizan podrían ser aplicables a otros tipos de evaluación, como la de servicios, la del rendimiento o la de centros. Al hilo de esta cuestión, hemos de señalar que ciertos autores han intentado establecer una distinción entre los conceptos de *programa* y *proyecto*. Así, por ejemplo, para el *Joint Committee on Standards for Educational Evaluation*, el término "programa" hace referencia al conjunto de actividades perdurables en el tiempo y a través de las cuales se presta un servicio. Por el contrario, los "proyectos" son conjuntos de actividades que tienen una duración determinada y que se realizan para un propósito concreto (1994). Otros autores, entre los que cabe destacar a Cohen y Franco (1993), consideran que los proyectos son aquellos programas que tienen contenido social. Sin embargo, desde nuestro punto de vista y puesto que en la práctica las evaluaciones de programas y proyectos se desarrollan de manera similar, las diferencias entre uno y otro término, en caso de existir, no afectan sustancialmente a las consideraciones que aquí realizamos. En cualquier caso, debe tenerse en cuenta que la naturaleza del objeto evaluado condiciona sustancialmente el proceso de evaluación, aunque existan algunas de las afirmaciones realizadas a propósito de un tipo de evaluación puedan ser aplicables a otro.

³ Sirvan de ejemplo las metaevaluaciones realizadas por House *et al.* (1977) y Stake (1986).

⁴ Sobre las premisas en las que descansa el paradigma positivista/científista véase Hammersley (1995: 2 y ss.). Una de las más brillantes críticas a este paradigma se encuentra en Guba y Lincoln (1989).

⁵ La finalidad de los diseños experimentales es determinar si los efectos producidos a raíz de la implantación de un programa son o no atribuibles al mismo. El experimento consiste en la formación, mediante procedimientos aleatorios, de uno o varios grupos de intervención (los que se someten al programa) y uno o varios de control (aquellos que no son beneficiarios) que son equivalentes en sus características básicas. En los diseños cuasiexperimentales los grupos de intervención y de control no son equivalentes, puesto que para su formación no resulta posible aplicar procedimientos aleatorios. Para mayor abundamiento véanse Campbell y Stanley (1963) y Cook y Campbell (1979).

a las cuales atribuyeron valor en función de una escala en la que la máxima valoración corresponde con el nivel máximo de calidad. Desde su perspectiva, sólo los diseños experimentales o cuasi-experimentales, el muestreo aleatorio, las técnicas cuantitativas de análisis de datos y los procedimientos multivariados merecen la máxima calificación⁶.

Por el contrario, las aportaciones de los autores de los enfoques *pluralistas/naturalistas*⁷ o de *cuarta generación*, por emplear la denominación acuñada por Guba y Lincoln (1989), tienden (i) a poner el énfasis en las cuestiones éticas y en el carácter intrínsecamente político de los procesos de evaluación y (ii) a destacar la excesiva simplicidad de las argumentaciones cuantitativas, que identifican "calidad" con utilización de técnicas cuantitativas (aunque suelen decantarse por los métodos de naturaleza cualitativa).

3. Algunas propuestas eclécticas para la evaluación de las evaluaciones

El debate entre cuantitativistas y cualitativistas en evaluación -que no es sino la reproducción en el ámbito evaluativo del debate más amplio que ha cruzado las Ciencias Sociales en los últimos años⁸, no ha sido estéril. De entrada, ha servido para advertir las ventajas e inconvenientes de ambos paradigmas y para constatar la dificultad de hallar unos criterios o estándares universalmente válidos que guíen al metaevaluador en su análisis crítico de la evaluación⁹. En función de ello, se tiende a aceptar mayoritariamente que dichos criterios habrán de ser adoptados en cada caso y sustentarse en el consenso al que sobre ellos lleguen los actores afectados a través de un proceso de negociación, concepto éste central en el enfoque naturalista¹⁰. En la actualidad se admite que el evaluador debe conocer y dominar todo tipo de técnicas y que la elección de unas u otras debe

hacerse en función de la naturaleza del objeto de estudio y de los objetivos de la investigación¹¹.

Precisamente, la polémica entre cuantitativistas y cualitativistas se ha enriquecido con contribuciones que, desde una perspectiva tendencialmente ecléctica, intentan proponer un esquema de análisis, generalizable en la medida de lo posible, que oriente al metaevaluador en el desempeño de su trabajo. Estas propuestas pretenden identificar las cuestiones que son de imprescindible tratamiento en cualquier metaevaluación.

De entre todas estas contribuciones hemos de destacar, en primer lugar, la obra de Michael Scriven (1991). Para este autor, el metaevaluador puede 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 *items*, 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* o sectores afectados), 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 no debe ser considerada como una lista de *desiderata* sino más bien como un catálogo de cuestiones que deben ser analizadas en el proceso de metaevaluación. Dichas cuestiones se refieren al diseño del estudio, el personal de la evaluación, la procedencia y la validez de los criterios empleados incluso para la elección del metaevaluador, los antecedentes y el contexto de la evaluación, su calendario, la identificación de las *audiencias*, 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 ésta y, por último, la significación global de la investigación evaluativa. Asimismo, en el KEC se contempla la posibilidad de que en el informe el metaevaluador incluya recomendaciones sobre el uso o el destino de la evaluación.

Otra de las propuestas que podríamos considerar ecléctica y que ha gozado también de gran predicamento

⁶ En la línea de esta argumentación se inscribirían también las obras de Campbell y Stanley (1963), Schuman (1967), Rossi y Williams (1972) y Wholey *et al.* (1970).

⁷ En puridad, el término *naturalismo* hace referencia al modelo de evaluación propuesto por Guba y Lincoln (1981) aunque, por extensión, se emplea también para designar a las corrientes pluralistas-cualitativistas.

⁸ Esta polémica se remonta a la década de los años veinte y treinta. La revalorización de la perspectiva cualitativa, que no se produjo hasta los años setenta, no se vio acompañada, como recuerda Bericat (1998), de un retroceso en el predominio de la orientación cuantitativa. Sabedores de su primacía los cuantitativistas han observado "casi con indiferencia" las críticas que les dirigían los cualitativistas. Como afirma Bericat, éstos "seguros de sí mismos, siguieron realizando su labor, perfeccionando sus técnicas, sin apenas levantar la cabeza e, incluso hasta muy recientemente, sin apenas volver la vista con la intención de asimilar los beneficios que la emergencia del cualitativismo pudiera reportarles" (1998: 27).

⁹ Así lo constata Patton (1986).

¹⁰ Pese a ello, Patton (1986) considera que cualquier evaluación puede ser juzgada según unos criterios de referencia de carácter general, tales como su conveniencia, su utilidad, su grado de factibilidad, su credibilidad y su relevancia.

¹¹ Patton señala que la utilización conjunta de distinto tipo de técnicas ha producido resultados muy interesantes en el campo de la evaluación, como muestran sus propias aportaciones (1982) o las de autores como Fetterman (1984), Reichardt y Cook (1979), Maxwell *et al.* (1985) o Firestone y Herriot (1984).

ha sido la realizada por el *Joint Committee on Standards for Educational Evaluation*.

En 1974, un comité nombrado por la *American Educational Research Association*, la *American Psychological Association* y el *National Council on Measurement on Education* y presidido por D.L. Stufflebeam¹², sugirió a estas organizaciones la elaboración de un conjunto de estándares con los cuales valorar la práctica de la evaluación educativa. El trabajo del comité -al cual se sumaron nueve asociaciones profesionales más- fue arduo y complejo, debido a las importantes divergencias existentes entre sus miembros acerca de cuestiones como la utilización de tests estandarizados o la inclusión de estándares relativos al personal¹³ o a organizaciones de evaluación. El *Joint Committee* recogió los resultados de sus reuniones en la obra *Estándares para la evaluación de programas, proyectos y materiales educativos*, publicada en 1981. Estos estándares fueron sometidos a un proceso de contrastación y revisión por parte de especialistas en el campo de la evaluación y de las Ciencias Sociales, así como por profesionales de la educación, antes de que su utilización fuera recomendada.

En 1989, al tiempo que se procedía a la acreditación oficial de los estándares por el *American National Standards Institute*, se comenzó a percibir la necesidad de revisar su contenido y de ampliar su ámbito de aplicación a programas médicos, asistenciales, militares o sociales. Se llegó así a una nueva y más precisa formulación de los estándares, en la que éstos son considerados como aspectos sustantivos de los cuatro rasgos característicos de la evaluación: *utilidad, viabilidad, pertinencia o conveniencia y precisión*¹⁴.

4. La metaevaluación desde una perspectiva cualitativa: la defensa de la ética, el pluralismo y la democracia de los procesos evaluativos

Como señalamos anteriormente, han sido los autores cualitativistas quienes han contribuido en mayor medida al desarrollo de la metaevaluación y quienes han insistido en que ésta ha de partir siempre de una premisa: el carácter intrínsecamente político de las evaluaciones. Pues bien, son varias las razones por las que, a juicio de los pluralistas-cualitativistas, se puede afirmar la naturaleza política de la investigación evaluativa¹⁵:

- (i) 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*¹⁶. La dinámica, los rasgos del contexto político en el que surge y los procedimientos, las estructuras y los actores que conforman este contexto inciden notablemente en su realización.
- (ii) La evaluación, en tanto sea requerida y/o financiada por los poderes públicos que han puesto en marcha una determinada política, está sujeta a la influencia y a la presión de dichos poderes.
- (iii) 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 (el respaldo que proporcionan los informes de expertos en una determinada materia), 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.
- (iv) La evaluación constituye además una actividad intrínsecamente política, desde el momento en el que en ella se ven involucrados intereses contrapuestos, políticamente relevantes, que luchan por obtener una mayor cuota de poder e influencia en la negociación dentro del proceso evaluativo. Los criterios conforme a los cuales se juzga la política, el diseño mismo de la evaluación no dejan de tener naturaleza política, por cuanto implican adoptar un posicionamiento u otro frente a cuestiones tan controvertidas como el acceso de determinados grupos a la evaluación, la propiedad de la información producida, etc...¹⁷.
- (v) 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.

¹² En la definición de Stufflebeam, la *metaevaluación* es considerada como "el proceso de diseño, obtención y empleo de información descriptiva y valorativa acerca de la utilidad, la viabilidad, la ética y el rigor técnico de una evaluación con el propósito de orientarla y dar cuenta públicamente de sus puntos fuertes y de sus puntos débiles" (1981: 151).

¹³ En 1988, el *Joint Committee* procedió a la publicación de un catálogo de estándares destinado a la evaluación del personal, *The Personnel Evaluation Standards*.

¹⁴ Los términos en inglés son, respectivamente, *utility, feasibility, propriety and accuracy*.

¹⁵ Puede resultar útil aplicar a las evaluaciones la imagen de las organizaciones como sistemas políticos propuesta por Morgan (1993: 153-213). Según esta imagen, las organizaciones son intrínsecamente políticas desde el momento en que en ella subsisten racionalidades e intereses contrapuestos.

¹⁶ Véanse Weiss (1987) y Palumbo (1987).

¹⁷ Hace ya años que algunos autores como (Weiss 1987) y (1991) y Norris 1990) denunciaron la situación en la que se encuentra la investigación evaluativa financiada con fondos públicos, su falta de credibilidad ante la injerencia cada vez mayor del gobierno en la elaboración de la agenda y el contrato de evaluación, esencialmente estableciendo, cuando no imponiendo, restricciones al uso y la difusión de la información. La *British Educational Research Association* dirigió también duras críticas contra los contratos de investigación que el gobierno imponía a los investigadores en un artículo titulado "Pay the Piper, Call the Tune", que fue publicado el 9 septiembre de 1993 en el diario británico *The Independence*.

Son pues especialmente proclives a ser instrumentalizadas, utilizadas para fines políticos o partidistas¹⁸.

Paralelamente, la constatación de la naturaleza y las implicaciones políticas de la evaluación ha llevado a los representantes de esta corriente a destacar la importancia de los principios de *participación* y *pluralismo*. Desde su punto de vista, los programas públicos, en tanto que productos de la actuación gubernamental, deben poder ser sometidos al juicio de todos los ciudadanos, especialmente al de quienes se ven directamente afectados por ellos. Su evaluación debe pues incorporar, en condiciones de igualdad, la pluralidad de opiniones existentes. La justicia y la ética del proceso evaluativo, el reconocimiento y el respeto a la pluralidad o la participación como medio para fomentar la utilización pasan, así, a convertirse en los estándares con los que determinar el valor de la evaluación. La metaevaluación habrá de centrarse, de este modo, en valorar el grado en que la investigación evaluativa cumple con dichos principios.

House es uno de los autores que más ha insistido en la necesidad de que la evaluación sea una experiencia pluralista y de que se atenga a requerimientos de carácter ético. El concepto de justicia merece en sus obras una especial consideración¹⁹.

Para este autor, una evaluación sólo es justa si en ella se respetan cuatro principios básicos: la *participación*, en condiciones de igualdad, de los distintos sectores afectados por el programa, la *autonomía* de estos sectores a la hora de valorar el programa (esto es, la ausencia de coerción), la *imparcialidad* o representación de todos los afectados, sin exclusión de ninguno y la *reciprocidad*, que se refiere al idéntico tratamiento que deben recibir todos los participantes en el proceso de evaluación.

Pero House avanza aún más en sus planteamientos al señalar que la evaluación debe convertirse en un medio de compensación de las desigualdades sociales. A su juicio, el evaluador ha de ocuparse de garantizar la presencia de los individuos o colectivos que no tengan un fácil acceso a la evaluación, acudiendo incluso a la introducción de mecanismos de discriminación positiva.

La ética y la justicia de los procesos evaluativos dependen, a su vez, de la participación de todos aquellos que se ven afectados por el programa. En esta línea argumentativa encontramos la obra de Robert Stake. En su opinión, para que la evaluación sea "sensible" a los

deseos de los distintos destinatarios²⁰, éstos han de participar en ella no sólo como informantes sino como decisores a lo largo de todo el proceso, desde el diseño hasta la elaboración del informe final.

Según otros autores, como Patton (1986) o Riquier (1997), la importancia de la participación de los *stakeholders*²¹ en la toma de decisiones radica en que potencia las posibilidades de utilización de las evaluaciones²². El concepto de *utilización* -que no se refiere únicamente a los resultados sino también al propio proceso de evaluación- está estrechamente vinculado al de *utilidad*. A juicio de Patton, *sólo se utilizará (o tendrá mayores probabilidades de utilizarse) aquella información que resulte relevante, significativa y que sintonice con los intereses de los stakeholders*²³. A su vez, como el concepto de utilidad varía de unos actores a otros es imprescindible que éstos tengan la oportunidad de participar en el proceso de evaluación y de hacer valer su idea de utilidad.

La utilización es pues, a juicio de Patton, una de las claves de la calidad de la evaluación y en ella incide decisivamente el llamado *factor personal*, que

²⁰ Este autor, en su estudio de la evaluación del programa "Cities-in-Schools" (1986), aplica el enfoque que él mismo ha dado en llamar "responsive evaluation". Aun cuando la traducción exacta del término "responsive" al castellano es la de "sensible", habitualmente se emplea la expresión "evaluación responsiva" o "respondente" para referirse a este modelo.

²¹ Optamos por el emplear el término en inglés porque su uso está ampliamente difundido entre los especialistas en evaluación. Dicho término, según recuerda Mendelow (1987; cit. por Patton 1997: 41), procede del mundo de la consultoría y se refiere a las personas que no poseen acciones o intereses directos en una compañía pero sin cuyo apoyo ésta dejaría de existir. Recordemos, asimismo que, según la definición realizada por el *Joint Committee* (1994: 25) dicho término se refiere a todos aquellos actores que están implicados en un proceso de evaluación o pueden verse afectados por el mismo.

²² El tema de la utilización de las evaluaciones es, sin duda, uno de los que mayor interés han despertado tanto entre los teóricos de la evaluación como entre los decisores y los responsables de los programas. Tal interés es consecuencia de la constatación de la "falta de uso de los resultados de las evaluaciones (que) ha llegado casi a alcanzar las proporciones de un escándalo nacional". Cualesquiera que sean las causas de la infrautilización, lo cierto es que los impactos de las evaluaciones suelen ser mínimos. Es extremadamente difícil encontrar un programa cuya evaluación haya supuesto grandes cambios. Cronbach señala que "Las evaluaciones tienen una influencia considerable sobre las políticas, pero esta influencia es 'indirecta', 'a largo plazo', 'secundaria' y 'conceptual' (1985). Como han destacado brillantemente Argyris (1982, 1994) y Argyris y Schön (1978), las organizaciones tienen grandes dificultades para aplicar el conocimiento adquirido a la resolución de problemas. Esta dificultad se acentúa, en el caso de los procesos evaluativos, por sus propias peculiaridades lo que hace que la utilización de sus resultados sea bastante improbable. Como apunta Chelmsky, si la evaluación cuestiona los supuestos y los principios del programa, es probable que sus resultados no lleguen a utilizarse porque lo contrario implicaría que la organización renunciaría a sus premisas y sus estrategias o incluso al mantenimiento de su *statu quo* (1995).

²³ Muy similar es la postura adoptada por Guba y Lincoln en su ya clásico *Effective Evaluation*. Estos autores señalan que la escasa utilización de las evaluaciones se debe esencialmente a la poca atención que los enfoques tradicionales han prestado a los intereses y necesidades informativas de las audiencias implicadas: "Con frecuencia esta falta se atribuye a la ignorancia, la pereza y la actitud esquiva de los decisores. Sin embargo, nos inclinamos a pensar que tal falta denota simplemente la pobreza de las evaluaciones tradicionales, que fallan porque no se ocupan de los intereses y las cuestiones que preocupan a las distintas audiencias y porque producen información que, aunque significativa desde el punto de vista estadístico, no genera un conocimiento verdaderamente valioso", Guba y Lincoln (1981: ix). Sobre el tema de la utilización pueden consultarse Alkin (1985) y Burry (1984).

¹⁸ Este es uno de los peligros contra el que nos advierte Morgan: "Comenzamos a ver política por todas partes y a buscar 'agendas escondidas' donde no hay agendas. por este motivo, la metáfora debe ser utilizada con precaución" (1993: 212). 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.

¹⁹ Ello se advierte tanto en sus reflexiones teóricas sobre evaluación (1980) como en sus trabajos de metaevaluación, entre los que destaca su crítica de la evaluación del programa "Follow Through" (House *et al.* 1977).

“representa el liderazgo, el interés, el entusiasmo, la determinación, el compromiso, la actitud positiva²⁴ y el apoyo de determinados individuos²⁵”. La utilización depende en gran medida de la voluntad y la actitud de los actores implicados. Así, “(...) cuando el factor personal está presente, cuando ciertas personas asumen la responsabilidad directa y personal de proporcionar los resultados a las personas adecuadas, las evaluaciones producen un impacto. Cuando este factor personal no aparece, hay una absoluta ausencia de impactos²⁶”.

Para otros autores la participación de los *stakeholders* en el proceso evaluativo es importante en la medida en que contribuye al fortalecimiento de la democracia.

En efecto, la evaluación *democrática* de la que habla Barry McDonald (1977), tiene también en el concepto de participación su principal referente. Para este autor, la evaluación democrática es aquélla que reconoce y representa la pluralidad de opiniones existentes sobre un programa y se concibe como un servicio de información a la sociedad sobre su rendimiento. Por contraposición, las evaluaciones *burocrática* y *autocrática* se orientan exclusivamente a la satisfacción de los intereses de la administración educativa. En la evaluación burocrática, el evaluador acepta sin condiciones los criterios de valor impuestos por las autoridades mientras que, en el caso de la evaluación autocrática, supedita su labor de legitimador del programa a que los decisores admitan sus recomendaciones.

La idea de participación cobra aún mayor relevancia en la obra de Kushner (2000). La evaluación se concibe como una oportunidad para someter a juicio a quienes están en el poder: “Puesto que cada programa social o educativo puede ser considerado como una reafirmación del contrato social (esto es, de las bases del poder, la autoridad, la estructura social, etc..) cada evaluación del programa constituye una oportunidad para revisar sus premisas y sus consecuencias²⁷”. En este sentido, la evaluación ha de hacerse más personal²⁸ y servir para aprehender las opiniones de quienes son objeto de los programas, especialmente de los más desfavorecidos y para informar sobre ellas a la sociedad. Sus puntos de vista deben prevalecer sobre los juicios de los expertos. La idea, en definitiva, es que la evaluación se construya por y para las personas, que éstas recuperen el protagonismo del proceso de evaluación. Ello contribuirá, sin duda, al fortalecimiento de la democracia.

En las reflexiones más recientes de House y Howe

(2000) la evaluación se considera también como pieza esencial de la democracia. Estos autores consideran que, en torno a la evaluación, han de establecerse de manera explícita una serie de principios democráticos que impidan que sólo los más poderosos tengan acceso a ella. Sólo sobre la base de estos principios –la inclusión, el diálogo y la deliberación- puede ser la evaluación una empresa fiable e íntegra que, al tiempo, favorezca el desarrollo de la democracia²⁹.

En una línea muy similar encontramos el modelo de *empowerment evaluation*, de Fetterman, Kaftarian y Wandersman (1996) aunque, en este caso, el objetivo de fortalecimiento de la democracia aparece de una forma más indirecta e implícita. Este modelo pretende proporcionar a las personas los conocimientos y herramientas necesarios para la mejora de sus programas, poniendo en marcha sus propios procesos de evaluación. El proceso de evaluación es así necesariamente participativo y democrático. El papel que en él asume el evaluador no es el clásico del juez experto sino el de un colaborador que ayuda a los implicados en el programa a evaluarse a sí mismos y a evaluar el programa.

Los últimos trabajos de Patton³⁰, establecen también un estrecho vínculo entre las ideas de democracia y utilización. En opinión de este autor, el aprendizaje que ofrece el proceso de evaluación a quienes participan en él puede resultar enriquecedor para la democracia, puesto que la solidez de ésta depende en gran medida del grado de información de la ciudadanía.

En definitiva, las corrientes naturalistas entienden, pues, que la evaluación ha de ser políticamente representativa. La argumentación que Pitkin realizara en su ya clásica obra, *El concepto de representación*, parece pensada para apoyar las tesis de esas corrientes. Señala esta autora que el ámbito político -y la evaluación está especialmente impregnada de connotaciones políticas- es precisamente “la clase de contexto en el que se hace relevante la representación como actividad sustantiva. Pues la representación no es necesaria allí donde esperamos soluciones científicamente verdaderas, allí donde no están involucrados compromisos de valor, decisiones ni juicios (...) Necesitamos la representación precisamente allí donde no nos contentamos con abandonar las cuestiones en manos del experto; podemos tener una representación sustantiva sólo allí donde el interés se vea implicado³¹”.

²⁴ El término utilizado en el original es *assertiveness*.

²⁵ Patton (1997: 44).

²⁶ Patton (1997: 47). Cronbach, en una línea argumentativa similar señala: “Nada tiene mayor importancia en el uso de las evaluaciones que el factor personal -el interés de los funcionarios por aprender de las evaluaciones y el deseo del evaluador por llamar la atención sobre lo que él sabe” (1985: 6).

²⁷ (2000: 32 y 33).

²⁸ El término empleado por el autor, que además da título a su obra, es *personalizing evaluation*.

²⁹ La idea del diálogo como fuerza y foro para la democratización ha sido también destacada por Greene (2001) y Tyan y DeStefano (2001). Para estos autores, el diálogo en la evaluación constituye un modo de mejorar la calidad democrática de los procesos decisorios al propiciar la implicación de las distintas perspectivas existentes sobre un programa.

³⁰ En especial el artículo publicado en 1998 y la ponencia presentada en la 4ª conferencia de la *European Evaluation Society* celebrada en Lausanne en octubre de 2000.

³¹ Pitkin (1985: 236).

5. Algunas cuestiones pendientes del cualitativismo³²

La evaluación ha llegado a ser considerada, hoy en día, como una actividad poco menos que imprescindible en la puesta en marcha de cualquier política o programa público. La tendencia, además, es que la importancia de la evaluación se vea reforzada en el futuro.

La metaevaluación, en tanto que análisis valorativo de las evaluaciones, adquiere su verdadero sentido en un contexto de consolidación de la evaluación de programas. Desde que Michael Scriven acuñara el término a finales de los años sesenta, la *metaevaluación* ha ido perfilándose como un tipo de investigación de indudable interés para determinar la calidad de las evaluaciones. Calidad que, como apuntan los autores pluralistas -que son quienes han contribuido en mayor medida al desarrollo de la metaevaluación- no debe medirse exclusivamente en función del rigor técnico de la evaluación sino también a partir de criterios tales como la utilidad de los resultados, la actitud de los poderes públicos con respecto a la propiedad de los estudios, el grado de pluralismo del proceso evaluativo o la medida en que éste pueda contribuir al fortalecimiento de la democracia

Como se puede, pues, advertir las propuestas de los representantes de las corrientes naturalistas, son claramente sugerentes. No obstante, son varias las cuestiones que permanecen abiertas:

- (i) La constatación de la naturaleza política de la evaluación por parte de las corrientes naturalistas (cuyos máximos representantes proceden del campo de la Psicología o la Pedagogía), no se ha visto acompañada de un análisis riguroso. El estudio de las evaluaciones ha de hacerse partiendo de unas categorías conceptuales y de unas herramientas metodológicas apropiadas, que permitan aprehender y explicar el carácter eminentemente político del proceso evaluativo, carácter que se deriva esencialmente de la presencia de actores, individuales y colectivos, con distintas estrategias y motivaciones. La Ciencia Política y, en particular, el enfoque del *policy analysis*³³, por su interés en el estudio del comportamiento de los actores, resultan ser, en este sentido, una inexcusable referencia para el metaevaluador. Además, el estudio de la política evaluativa de los gobiernos puede arrojar alguna luz sobre una cuestión trascendental: el control que puedan ejercer los poderes públicos sobre los resultados de estas investigaciones³⁴.

³² La necesidad de limitar la extensión de este artículo nos obliga a dejar tan sólo apuntadas estas cuestiones. El esquema de análisis aplicable al estudio de las relaciones entre actores en la metaevaluación puede encontrarse en García Sánchez (2003).

³³ Muchos son los autores que han subrayado la importancia de los actores (incluido el Estado) en las políticas públicas. Merece la pena resaltar, en este sentido, las obras de Crozier y Friedberg (1977), Hecló y Wildavsky (1974), Jobert y Muller (1988), Dupuy y Thoenig (1983), Regonini (1990-91), Padioleau (1982) o Mèny y Thoenig (1992).

³⁴ Creemos que la conocida afirmación de Ashford (1976) sobre las políticas es perfectamente aplicable a las evaluaciones: así, consideramos a éstas como una forma de descripción del comportamiento institucional.

- (ii) El tema de la participación de los actores en los procesos de evaluación no plantea, como sugieren algunos autores pluralistas, problema ético sino metodológico. La evaluación es, ante todo, investigación y, como tal está sujeta a ciertas exigencias de carácter técnico. Por ello, la tarea primordial del metaevaluador es determinar y valorar la *capacidad explicativa* de la evaluación. El rigor que debe presidir cualquier investigación invalida la propuesta pluralista de que se introduzcan en la evaluación mecanismos de discriminación positiva, que supongan sobrestimar la importancia de unos grupos en detrimento de otros.

- (iii) Pero, por otro lado, si se acepta que en una democracia los ciudadanos tienen un derecho legítimo a valorar la actuación de los poderes públicos, el problema alcanza una dimensión que excede el ámbito de lo metodológico. ¿Es, por tanto, admisible excluir a ciertos actores del proceso de evaluación? ¿Ha de ocuparse el metaevaluador de comprobar si los poderes públicos han obstaculizado su participación? En muchas ocasiones, se entiende imprescindible que la metaevaluación no se limite a un mero examen de la calidad técnica de una evaluación sino que se ocupe también de analizar de qué modo intervienen los actores y cuáles son las relaciones de poder que se establecen en el seno del proceso evaluativo³⁵. El fin último de la metaevaluación sería pues, determinar si la evaluación (entiéndase aquella que se financia con fondos públicos) cumple unos mínimos contenidos democráticos.

Prof. Ester GARCIA SÁNCHEZ

Referencias bibliográficas

- Alkin, M.C. (1985), *A Guide for Evaluation Decision Makers*, Beverly Hills, CA: Sage.
- Argyris, C. (1982), *Reasoning, Learning and Action*, San Francisco: Jossey-Bass.
- (1994), "Good Communication that Blocks Learning", *Harvard Business Review*, vol. 72, núm. 4.
- Argyris, C. y Schön, D.A. (1978), *Organizational Learning: A Theory of Action Perspective*, Reading, Mass.: Addison-Wesley.
- Ashford, D. (1976), *Democracy, Decentralization and Decision in Sub-National Politics*, Beverly Hills: Sage.
- Bericat, E. (1998), *La integración de los métodos cuantitativo y cualitativo en la investigación social*, Barcelona: Ariel

³⁵ En primer lugar y, como se muestra en García Sánchez (2003), sería necesario proceder a la identificación de los actores implicados, determinar de qué modo intervienen en el mismo y el tipo de decisiones en las que participan así como el rol que desempeñan en la evaluación.

- Bernstein, I. y Freeman, H.E. (1975), *Academic and Entrepreneurial Research: Consequences of Diversity in Federal Evaluation Studies*, Nueva York: Russell Sage.
- Burry, J. (1984), "Synthesis of the Evaluation Use Literature", *Evaluation Productivity Project*, Center for the Study of Evaluation, Universidad de California-Los Angeles.
- Campbell, D.T. y Stanley, J.C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally.
- Chelimsky, E. (1995), "New Dimensions in Evaluation" en R.C. Rist y R. Picciotto, eds., *Proceedings of the 1994 World Bank Conference*.
- Cohen, E. y Franco, R. (1993), *Evaluación de proyectos sociales*, Madrid: Siglo XXI.
- Cook, T.D. y Campbell, D.J., (1979), *Quasi-experimentation. Design, Analysis. Issues for Field Experimentation*, Cambridge, Ma: Houghton Mifflin Co.
- Cronbach, L.J. et al., (1985), *Toward Reform of Program Evaluation*, San Francisco: Jossey-Bass.
- Crozier, M. y Friedberg, E. (1977), *L'acteur et le système*, París: Seuil.
- Dupuy, F. y Thoenig, J-CI. (1983), *Sociologie de l'Administration française*, París: A. Collin.
- Fetterman, D.M. (1984), "Ethnography in Educational Research: The Dynamics of Diffusion" en D.M. Fetterman, ed., *Ethnography in Educational Evaluation*, Beverly Hills, CA: Sage.
- Kaftarian, S.J. y Wandersman, A., eds. (1996), *Empowerment Evaluation. Knowledge and Tools for Self-Assessment & Accountability*, Thousand Oaks, CA: Sage.
- Firestone, W.A. y Herriot, R.E. (1984), "Multisite Qualitative Policy Research: Some Design and Implementation Issues" en D.M. Fetterman, *Ethnography in Educational Evaluation*, Beverly Hills, CA: Sage.
- García Sánchez, E. (2003), *La evaluación de programas de reforma educativa en España. Tres estudios de caso desde un enfoque de metaevaluación*, tesis doctoral, Universidad Complutense de Madrid.
- Glass, G.V. y Worthen, B.R. (1971), "Evaluation and Research: similarities and differences", *Curriculum Theory Network*, otoño.
- 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.
- (1989), *Fourth Generation Evaluation*, Newbury Park, CA: Sage.
- Greene, J.C. (2001), "Dialogue in Evaluation: A Relational Perspective", *Evaluation*, vol. 7, núm. 2.
- Hammersley, M. (1995), *The Politics of Social Research*, Londres: Sage.
- Hecl, H. y Widavsky, A. (1974), *The Private Government of Public Money. Community and Policy Inside British Politics*, Londres: McMillan.
- House, E.R. et al. (1977), *No Simple Answer. Critique of the Follow Through Evaluation*, (Urbana: CIRCE/ University of Illinois).
- House, E.R. (1980), *Evaluating with Validity*, Beverly Hills, CA: Sage.
- (1993), *Professional Evaluation. Social Impact and Political Consequences*, Newbury Park, CA: Sage.
- House, E.R. y Howe, K.R. (2000), "Deliberative Democratic Evaluation", *New Directions for Evaluation*, 85.
- Jobert, B. y Muller, P. (1988), *L'Etat en action*, París: PUF.
- Joint Committee on Standards for Educational Evaluation (1988), *The Personnel Evaluation Standards*, Thousand Oaks: Sage.
- Joint Committee on Standards for Educational Evaluation (1994), *The Program Evaluation Standards*, Thousand Oaks: Sage.
- Kushner, S. (2000), *Personalizing Evaluation*, Londres: Sage.
- Lessinger, L.M. (1970), *Every Kid a Winner: Accountability in Education*, New York: Simon and Schuster.
- MacDonald, B. (1977), "A political classification of evaluation studies" en D. Hamilton, ed., *Beyond the Numbers Game*, Londres: MacMillan.
- Maxwell, J.A. et al. (1985), "Combining Ethnographic and Experimental Methods in Educational Research: A Case Study", en M.D. Fetterman y M.A. Pitman, *Beyond the Status Quo: Theory, Politics and Practice in Ethnographic Evaluation*, Washington, DC: Cato Institute.
- Mendelow, A.L. (1987), "Stakeholder Analysis for Strategic Planning and Implementation" en W.R. King y D.L. Cleland, eds., *Strategic Planning and Management Handbook*, Nueva York: Van Nostrand Reinhold.
- Mèny, Y. y Thoenig, J-C. (1992), *Las políticas públicas*, Barcelona: Ariel.
- Morgan, G. (1993), "Intereses, conflictos y poder: las organizaciones como sistemas políticos" en C. Ramíó y X. Ballart, *Lecturas de teoría de la organización*, vol. II, Madrid: Ministerio para las Administraciones Públicas.
- Norris, N. (1990), *Understanding Educational Evaluation*, Londres/Norwich: Kogan Page/Centre for Applied Research in Education.
- Padioleau, J. (1982), *L'Etat au concret*, París: PUF.
- Palumbo, D.J. (1987), *The Politics of Program Evaluation*, Beverly Hills, CA: Sage.
- Patton, M.Q. (1986), *Utilization-Focused Evaluation*, Thousand Oaks, CA: Sage.
- (1997), *Utilization-Focused Evaluation*, 3ª edición, Thousand Oaks, CA: Sage.
- (1998), "Discovering Process Use", *Evaluation*, vol. 4, núm. 2.
- (2000), "A Vision of Evaluation that Strengthens Democracy", ponencia presentada en la IV Conferencia de la Sociedad Europea de Evaluación, Lausana: 12-14 de octubre.
- Pitkin, H. (1985), *El concepto de representación*, Madrid: Centro de Estudios Constitucionales.

Mapamond

- Provus, M.M. (1971), *Discrepancy Evaluation*, Berkeley: McCutchan.
- Regonini, G. (1990-91), "El estudio de las políticas públicas" en *Documentación Administrativa*, núms. 224-225, número monográfico, octubre-marzo.
- Reichardt, C.S. y Cook, T.D. (1979), "Beyond Qualitative Versus Quantitative Methods" en C.S. Reichardt y T.D. Cook, eds., *Qualitative and Quantitative Methods*, Beverly Hills, CA: Sage.
- Riquier, G. (1997), "Stakeholders' Participation in Public Policy Evaluations: Impact on Organizational Learning", ponencia presentada en la II Conferencia de la Sociedad Europea de Evaluación, Estocolmo: 6-7 de marzo.
- Rossi, P.H. y Freeman, H.E. (1989), *Evaluación. Un enfoque sistemático para programas sociales*, México: Trillas.
- Rossi, P.H. y Williams, W. (1972), *The New Jersey Income Maintenance Experiment: Theory, Practice and Politics*, Nueva York: Academic Press.
- Ryan, K. y DeStefano, L. (2001), "Dialogue as a Democratizing Evaluation Method", *Evaluation*, vol. 7, núm. 4.
- Schuman, E.A. (1967), *Evaluative Research*, Nueva York: Russell Sage Foundation.
- Scriven, M. (1991), *Evaluation Thesaurus*, Newbury Park, CA: Sage.
- Seligman, R. (1973), "College Guidance Program", *Measurement and Evaluation in Guidance*, 6.
- Stake, R.E. (1986), *Quieting Reform. Social Science and Social Action in an Urban Youth Program*, Urbana/Chicago: University of Illinois Press.
- Stufflebeam, D.E. (1981), "Metaevaluation: Concepts, Standards and Uses" en R. Berk, ed., *Educational Evaluation: The State of the Art*, Baltimore: The Johns Hopkins University Press.
- y Shinkfield, A.J. (1993), *Evaluación sistemática. Guía teórica y práctica*, Madrid: Paidós/Ministerio de Educación y Ciencia.
- Weiss, C.H. (1987), "Where Politics and Evaluation Research Meet" en D.J. Palumbo, *The Politics of Program Evaluation*, Beverly Hills, CA: Sage.
- (1991), "Evaluation Research in the Political Context" en M.W. McLaughlin y D.C. Philips, eds., *Evaluation and Education: A Quarter Century*, Chicago: University of Chicago Press.
- (1998), *Evaluation*, New Jersey: Prentice Hall.

Mapamond
