

Revista mexicana de investigación educativa

COMIE

comie@servidor.unam.mx

ISSN (Versión impresa): 1405-6666

MÉXICO

2004

Miguel Ángel Santos Guerra / Tiburcio Moreno Olivos

¿EL MOMENTO DE LA METAEVALUACIÓN EDUCATIVA? CONSIDERACIONES
SOBRE EPISTEMOLOGÍA, MÉTODO, CONTROL Y FINALIDAD

Revista mexicana de investigación educativa, octubre-diciembre,
año/vol. IX, número 023

COMIE

Distrito Federal, México

pp. 913-931

¿EL MOMENTO DE LA META-EVALUACIÓN EDUCATIVA?

Consideraciones sobre epistemología, método, control y finalidad

MIGUEL ÁNGEL SANTOS GUERRA / TIBURCIO MORENO OLIVOS

Una buena parte de la discusión sobre los fenómenos educativos está construida con términos de una enorme complejidad semántica. La palabra educación encierra, en sí misma, tal polisemia que dentro de sus acepciones cabe todo tipo de interpretaciones y de actuaciones. Una persona pretende educar y por ello golpea al alumno o al hijo “para que aprenda”; otra entenderá que, para educar, no puede aconsejar por temor a la manipulación. Los dos piensan que “están educando”. Lo mismo debe decirse del concepto de evaluación. Evaluación se llama a la medición de resultados realizada mediante pruebas estandarizadas, a las auditorías, a la aplicación de pruebas objetivas, al análisis de procesos, a la investigación en la acción, etcétera. No digamos nada de las aplicaciones, utilidades, usos y abusos que se llevan a cabo con la evaluación (Santos Guerra, 1993). Su polisemia se multiplica hasta el infinito cuando pretendemos llegar a sus dimensiones más profundas.

Si eso sucede con los conceptos, ¿qué no pasará con los procesos que incluyen muchos de ellos? Cuando hablamos de evaluación hemos de precisar de qué evaluación se trata, a qué llamamos programa, qué es mejora de la práctica, qué entendemos por valores en la acción educativa, etcétera. De no hacerlo así, correremos el riesgo de estar hablando de fenómenos y procesos que nada tienen que ver con lo que otros piensan sobre ellos.

Miguel Ángel Santos es catedrático de Didáctica y organización escolar de la Universidad de Málaga, campus de Teatinos 29071, Málaga, España. CE: arbol@uma.es

Tiburcio Moreno es investigador del Centro de Investigación en Ciencias y Desarrollo de la Educación (CINCIDE) de la Universidad Autónoma del Estado de Hidalgo, México. CE: moreno@uaeh.reduaeh.mx

Es necesario respetar las reglas de la argumentación (Weston, 1994) para librar el discurso de trampas y sofismas. Resulta sorprendente analizar los abusos lógicos que llevan a conclusiones tajantes en el ámbito de la educación y de la evaluación. No tener en cuenta que muchos discursos, incluso los que pretenden basarse en las investigaciones y evaluaciones más ambiciosas, están larvados por intereses y perspectivas ideológicas y políticas, es engañarse a sí mismo o engañar a otros.

Muchos discursos sobre la educación y sobre la evaluación están trucados por la ambigüedad o la peculiaridad de las acepciones de las que se parte. La secuencia “educación, evaluación, metaevaluación” va haciendo más complejo y más difícil el entendimiento y más discutible la claridad. Es interesante, al respecto, el análisis de la lógica formal de una evaluación (y de su metaevaluación) realizada por House sobre la evaluación de Glass (1972) y replicada por Scriven (1976).

El principal problema surge cuando, por creer que todos estamos hablando de lo mismo, no aclaramos lo que decimos. La delimitación conceptual es uno de los pilares sobre los que se construye el pensamiento compartido. Otro es el rigor en la argumentación, frecuentemente cargada de sofismas (Santos Guerra, 1997) cuando se refiere a la educación:

Llevamos ya unos años haciendo evaluaciones de los centros. Es hora de iniciar un discurso sobre esos procesos. ¿Qué está pasando? ¿Cómo se está haciendo? ¿Qué se está consiguiendo? ¿Qué efectos secundarios está produciendo? No se trata de hacer evaluaciones sin ton ni son.

Como un metalenguaje, que es un lenguaje que se usa para hablar del lenguaje mismo, la metaevaluación consiste en analizar el proceso de evaluación. Digamos, sacrificando el estilo a la claridad, que metaevaluación es la evaluación de la evaluación. Aunque el prefijo “meta” puede tener connotaciones temporales (después de), no necesariamente se realiza la metaevaluación una vez finalizado el proceso de evaluación. Puede ser paralelo a la misma.

Existe una función retroalimentadora de la evaluación que ha sido frecuentemente olvidada, por eso se han repetido evaluaciones escasamente aprovechables para generar conocimiento y mejora y se han repetido errores de forma extrañamente irracional. Cuando se evalúa mucho y se mejora poco, algo está fallando en el proceso.

Las ideas que se aplican a la evaluación son trasladables a la metaevaluación. No es, pues, un proceso esencialmente técnico sino que tiene naturaleza política y ética. Desde la primera y exclusiva acepción, se trataría de un proceso mecánico, aparentemente riguroso, consistente en aplicar esquemas o principios y criterios de rigor, lo cual supondría despojar a la evaluación de sus componentes esenciales: a quién sirve, a quién trata de beneficiar.

Muchas preguntas que se formulan respecto de la evaluación han de ser planteadas en un metanálisis coherente, concienzudo y fundamentado. Se puede aplicar un tipo de metaevaluación a una evaluación que se ha basado en criterios radicalmente diferentes. Así, se verán reflejados los abusos, las contradicciones, los problemas y los desajustes.

No se trata, pues, de concluir, al estilo de las investigaciones experimentales, si existe o no validez, sino de avivar la reflexión para generar comprensión rica y relevante sobre los programas evaluados. Esto convierte a la evaluación no tanto en una tarea de expertos cuanto en un proceso democrático de discusión.

No basta interrogarse sobre la calidad de los datos que se recogen en el informe o sobre el rigor de los métodos que se han aplicado para extraer información, sino que es preciso preguntarse por la naturaleza de la iniciativa, por la finalidad del proceso, por las reglas éticas que lo han inspirado, por la utilización que se ha hecho de la información, por el control democrático de los informes, por las secuelas que la evaluación ha dejado. Plantearé seis contextos básicos de exploración que permiten analizar ideas en el marco concreto y temporal de su desarrollo. Aunque todos están interconectados, seguiré su cronología evolutiva.

1) Contexto de constitución

Estudia los mecanismos de generación de la evaluación, de surgimiento y negociación de todo el proceso. La historia de la génesis es muy interesante porque en ella pueden encontrarse los embriones de posteriores situaciones y actitudes. No es lo mismo una iniciativa surgida de la jerarquía e impuesta, que otra que parte de los protagonistas. No es lo mismo una iniciativa que procede de los directivos de una institución, que otra que parte de los integrantes de la comunidad.

En ese contexto cabe incluir todo el proceso de negociación. Con quién se hace, qué contenidos aborda, qué problemática surge y cómo se trata.

En la negociación inicial está el mapa genético de la evaluación. Hay que diferenciar negociaciones meramente formales, de otras que llegan hasta el fondo de las cuestiones. Hay que saber si se ha negociado lo insustancial dejando lo fundamental como incuestionable.

cuestiones

¿Quién y por qué ha tomado la iniciativa de hacer la evaluación? ¿Qué intereses (explícitos o implícitos) determinan la puesta en marcha de la evaluación? ¿Con quién se ha negociado la evaluación? ¿Ha sido una negociación real o meramente protocolaria y formal? ¿Ha estado la evaluación planteada como un mecanismo de poder o de control?

2) Contexto de justificación

Analiza los métodos elegidos para la obtención de los datos, su aplicación, la información a que han dado lugar y el tratamiento que se ha hecho de ella. Del rigor en la selección y aplicación de los métodos se deriva una buena parte del rigor de la evaluación.

Los procedimientos y estrategias utilizados para el desarrollo de la evaluación permiten conocer “el cómo” del proceso. La manera de proceder, la aceleración, el clima, las relaciones, las explicaciones que se ofrecen y que se piden constituyen un elemento importante del entramado de la evaluación. Aquí debe analizarse el tratamiento que se ha hecho de los datos, el rigor de la argumentación utilizada así como la triangulación –si se ha hecho– de la información procedente de distintos sujetos y de diferentes métodos.

Cuestiones

¿Qué tipo de métodos se ha utilizado para hacer la exploración? ¿Se ha adaptado la naturaleza y la secuencia de los métodos a las características de los interesados? ¿Se ha diversificado la metodología para recoger una información más rica y plural? ¿Se ha garantizado el anonimato a los informantes?

3) Contexto de presentación

Estudia, entre otros aspectos, el contenido y la forma de presentación del informe, la adecuación del mismo a las diversas audiencias, la estructura lógica, la coherencia interna, la extensión así como la claridad expositiva.

La elaboración del informe exige rigor en la argumentación que se realiza con los datos recogidos, de manera que no se produzcan silencios indebidos o conclusiones caprichosas. La forma de presentar la información ha de tener como finalidad fundamental la comprensión de los lectores, no la epifanía de la erudición del evaluador.

Cuestiones

¿Cómo se han trabajado los datos recogidos? ¿Se ha hecho triangulación de métodos, de sujetos, de tiempos? ¿Cómo se ha presentado la información? ¿Es inteligible el contenido de los informes? ¿Se han hecho informes diversos para diferentes audiencias?

4) Contexto de negociación

El informe ha de ser entregado a sus destinatarios de forma rápida, cómoda y directa. Al decir directa nos referimos a que tenga carácter democrático y no esté jerarquizada. No es bueno entregar el informe sólo al director de una institución o al responsable del programa. Las condiciones de la negociación han de ser razonables en cuanto al tiempo y a las formas de discusión. Las actitudes del evaluador permitirán descubrir si se trata de una negociación real o de un simulacro.

Cuestiones

¿Se han entregado los informes en las fechas convenidas? ¿Se ha concedido un tiempo suficiente para su atenta lectura? ¿Se ha negociado democráticamente? ¿Se ha podido hablar con libertad respecto de las cuestiones problemáticas? Si no se ha llegado al acuerdo en alguna cuestión, ¿se ha incorporado el contenido de la discrepancia de los participantes?

5) Contexto de valoración

Analiza las motivaciones, las intenciones, las actitudes, las presiones, los cambios y las vicisitudes que han manifestado los protagonistas durante el proceso de evaluación. En el contexto de valoración se tiene en cuenta el “cómo” se ha realizado todo el proceso. Por protagonistas me refiero, fundamentalmente, a los patrocinadores de la evaluación, los evaluadores y los evaluados. La evaluación puede tener diversas influencias nacidas de actitudes espurias, que es necesario poner de manifiesto, ya que pueden adulterar el sentido y el significado de todo el proceso.

Cuestiones

¿Se han producido cambios respecto del plan inicial? Si se ha producido algún conflicto, ¿cómo se ha afrontado y solucionado? ¿Han existido presiones por parte de algún grupo o persona para dirigir la evaluación? ¿Se ha producido algún sesgo en la utilización de la información?

6) Contexto de difusión

Una vez negociados los informes de la evaluación, ha de procederse a su difusión siguiendo unas reglas que deben estar consensuadas por los protagonistas. Es preciso tener en cuenta los derechos de los ciudadanos a conocer lo que sucede con la educación y, a la vez, el derecho de los profesionales a la privacidad en el ejercicio de su actividad. La demora excesiva, la adulteración de los informes, la selección de alguna de sus partes, la edulcorización de los contenidos, etcétera, son vicios que, si existen, la metaevaluación deberá descubrir y poner de manifiesto.

Cuestiones

¿Se ha hecho público el informe? ¿Se ha consensuado la publicación en todos los aspectos relevantes: contenido, forma, tiempo? ¿Han existido presiones del patrocinador o de los evaluados para eliminar alguna parte del informe? ¿El contenido del informe difundido ha sido claro, breve, conciso y pertinente?

La metaevaluación es una plataforma de discusión que une a patrocinadores, evaluadores, diversas audiencias y a la sociedad en general, que debe conocer no sólo los resultados de las evaluaciones realizadas sobre el funcionamiento de servicios públicos sino la metaevaluación que se realiza sobre ellas ya que ésta funciona como la garantía de la calidad de la evaluación. La metaevaluación, como se apunta en el título, tiene que ver con diferentes cuestiones teleológicas, que abordaremos a continuación.

El rigor de la evaluación

La metaevaluación se pregunta por el rigor. ¿Es fiable el conocimiento extraído de la evaluación o es fruto de la arbitrariedad, de los intereses y de la torpeza? Existe un problema que consiste en trasladar los términos y conceptos de la evaluación experimental a la de carácter cualitativo. La servidumbre del paradigma etnográfico ha hecho que se haya copiado y trasladado el modelo de validez interna.

“En sentido amplio –dice House (1994)– entiendo por validez la cualidad de merecer reconocimiento”. Ahora bien, cada modelo de evaluación hace una interpretación de lo que ha de entenderse por validez. House habla de ocho tipos o modelos de evaluación. Cuatro son calificados de “objetivistas” por su epistemología, y son los enfoques del análisis de sistemas, de objetivos, de decisión y la evaluación que prescinde de objetivos. En mayor o menor grado, todos basan su explicación de validez en métodos explícitos.

Las peculiaridades genéricas de estos cuatro modelos de evaluación podrían sintetizarse en las siguientes:

- 1) se derivan de la tradición teórica empírica de las ciencias sociales;
- 2) se manifiestan neutrales desde el punto de vista axiológico;
- 3) basan su pretensión de validez en métodos explícitos calificados de científicos;
- 4) se valora mucho en ellos la fiabilidad de la medida;
- 5) su fuerza está en el rigor del método prescrito, no en las interpretaciones; y
- 6) se valora mucho la posibilidad de confirmar las hipótesis.

Otros cuatro enfoques son calificados como “subjetivistas”. El método crítico, y los enfoques de revisión profesional, cuasijudicial y del estudio de casos. Éstas son algunas de las características fundamentales de los métodos subjetivistas en lo que se refiere a su validez:

- 1) fundan su presunta validez en la apelación a la experiencia y no al método científico;
- 2) suponen que el saber es, en gran medida, tácito más que explícito;
- 3) es muy importante la significación y la interpretación;
- 4) se asume que cada evaluador y cada lector son capaces de establecer conclusiones de forma individual;
- 5) lo que es válido para una persona puede no serlo para otra;
- 6) la evaluación ha de ser comprensible para los destinatarios;
- 7) están internamente ligados con la visión fenomenológica de la persona; y
- 8) la exactitud de la descripción y la utilidad para los destinatarios son conceptos clave.

Cada uno de los modelos tiene ciertas exigencias para garantizar la validez. La tarea del metaevaluador consiste en analizar todos estos presupuestos, es decir, que deberá tomar en cuenta el modelo de evaluación utilizado y, además, ir más allá para poner sobre el tapete sus limitaciones. La forma de pensar y de actuar del evaluador tienen que ir acompañadas de rigor en la manera de recoger la información al igual que de consistencia en el análisis de los datos. No basta una intención noble y sincera, ni un compromiso con los valores de la sociedad. Es necesario el rigor en la manera de acercarse a la realidad, de escuchar lo que “dice” y de interpretarlo de forma rigurosa y creíble. Existen, desde nuestra perspectiva, cinco modos de responder a esta cuestión, aunque no todos son de la misma naturaleza ni tienen la misma importancia para solucionar el problema epistemológico y metodológico:

Criterios de rigor

Es necesario garantizar el rigor de la evaluación, tanto en la cuantitativa como en la cualitativa, aunque el concepto de validez tenga una diferente acepción en un caso y otro.

El concepto de validez, de acuerdo con Hopkins (1989), consiste en que los evaluadores demuestren que las observaciones que registran y analizan corresponden a lo que pretenden registrar y analizar. La falta de validez se debe a errores sistemáticos. La fiabilidad consiste en evitar las fluctuaciones aleatorias en las observaciones que registran, la carencia de fiabilidad se produce a causa de errores aleatorios. Esencialmente, la fiabilidad está relacionada con la consistencia de los resultados y se refiere al requisito de que cualquier otro evaluador o el mismo en otra ocasión podrá replicar la evaluación original, consiguiendo datos comparables.

McCormick y James (1995) hablan de dos tipos de fiabilidad: la interjueces, que alude al grado de acuerdo entre dos evaluadores que registran sus observaciones del mismo fenómeno o de fenómenos iguales o similares, y la intrajueces que se refiere a la consistencia de las observaciones del mismo evaluador en distintas ocasiones.

Es evidente que cada una de estas definiciones se acomoda de forma distinta a la evaluación cuantitativa y a la cualitativa. Cuando hablamos de replicar una evaluación, está claro que nunca habrá dos situaciones idénticas si nos referimos a una exploración etnográfica. Cada experien-

cia es única, irrepetible, situada en un contexto peculiar, sometida a expectativas e influencias diversas, etcétera. Lo anterior no quiere decir que la evaluación cualitativa tenga menores exigencias o que sea menos rigurosa que la cuantitativa. Cada una tiene su peculiaridad y su problemática. Así como las pruebas de la validez parecen más claras y definitivas en la evaluación cuantitativa, no es menos cierto que hay que poner en cuestión el rigor de la medición en fenómenos difícilmente cuantificables y aislables.

La aparentemente más rigurosa forma de comprobar la validez en las evaluaciones cuantitativas, no es de ningún modo absoluta e indiscutible. Es exclusivamente estocástica, se basa en la probabilidad. Hemos de remitirnos a Guba (1981) para plantear esta cuestión ya que su artículo, alejado ya en el tiempo, sigue siendo un punto de referencia. Dice Guba que no existen criterios absolutos que garanticen el rigor, pero que podemos tener indicadores de su presencia. Una de las preocupaciones más importantes respecto de cualquier evaluación son los criterios de credibilidad. ¿Tienen calidad los datos, gozan de rigor las conclusiones? Sin estas preguntas, todo el edificio se viene abajo.

El contagio de la más arraigada y difundida investigación experimental ha pesado, y sigue haciéndolo, sobre las concepciones del diseño, del desarrollo y de los criterios de validez interna. Por eso se buscan criterios de parecida naturaleza (y a veces con la misma terminología) a los de la investigación cuantitativa. Es un error, ya que la tipología de los procesos es diferente.

No hay un criterio absoluto de credibilidad. Ninguno de ellos constituye una segura e inequívoca prueba de rigor. Lo cual no quiere decir que no sea necesario preguntarse por los indicadores que dan fiabilidad a la indagación.

Juicio crítico de expertos

El metaevaluador se dirige a los evaluadores, pero también al patrocinador y a las diversas audiencias de la evaluación. Es importante elegir cuidadosamente al metaevaluador, ya que su independencia así como las garantías de independencia que ofrece, son el fundamento de un trabajo riguroso.

Puede elegirse a una persona afín al paradigma de la evaluación realizada o bien a alguien que se halle situado en posiciones diferentes. Ten-

dría dudosa eficacia elegir a alguien poco sensible al modelo de evaluación con el que se ha trabajado. La elección puede realizarla el patrocinador en connivencia con los evaluadores o bien alguna de estas partes de forma independiente.

Es obvio que el metaevaluador realizará el trabajo desde la óptica de su peculiar visión de lo que ha de ser un proceso de evaluación. Debería explicitar en su informe cuáles son los presupuestos de los que parte. Además, ha de disponer de toda la documentación escrita de la evaluación: documentos de negociación, informes, cuadernos de campo, etcétera. Y puede entrevistar a las personas que hayan participado en la evaluación.

Conviene tener en cuenta que la evaluación es todo el proceso, no únicamente el informe final. Por eso es necesario seguir las pistas de revisión disponibles. O, mejor aún, comenzar el proceso de metaevaluación de forma paralela a la evaluación misma. El metaevaluador hace juicios, pero también plantea sugerencias y preguntas (Angulo, 1993). Ese tipo de enfoque ayuda a los evaluadores tanto descubrir las limitaciones como a mejorar su trabajo; pone un espejo (Santos Guerra, 1995) para que los evaluadores puedan ver con más precisión cuál ha sido su manera de proceder. Mediante ese espejo –que el metaevaluador (dada su condición de independiente y dado el tiempo de que dispone para hacer el trabajo) coloca de forma precisa– el evaluador puede verse de manera más rigurosa.

Opinión de los protagonistas

Los protagonistas de la evaluación tienen las claves del proceso. Saben si se han respetado las reglas, conocen cómo se han aplicado los métodos, quiénes han sido los agentes de la negociación de los informes, etcétera. Recabar por escrito o verbalmente su opinión sobre el proceso de evaluación es un modo excelente de comprobar el rigor y pueden realizarlas no sólo los evaluados sino también los evaluadores. Unos y otros contrastan esas opiniones que en algún caso se publican como complemento del informe. Las opiniones y las actitudes de los protagonistas pueden modificarse a lo largo del proceso de evaluación. Es interesante saber cómo y por qué se ha producido ese cambio.

Para que las opiniones sean sinceras es preciso crear un clima de libertad en el que se puede expresar sin ambages lo que se piensa. De lo con-

trario, no tendrán ningún valor o, lo que es peor, ofrecerán pistas falsas de análisis. Los protagonistas pueden plantear cuestiones relativas al proceso, al rigor de la aplicación de los métodos y de la utilización de los datos y, también, sobre los aprendizajes realizados.

No conviene distanciar mucho esta exploración a los protagonistas de la realización de la evaluación, si es que se efectúa una vez terminada. Digo esto porque la separación en el tiempo hace olvidar selectivamente partes de lo que ha sucedido. Pueden recabar esta opinión no sólo los evaluadores sino otras personas que tengan independencia respecto del proceso realizado. Ese hecho ofrecerá mayores garantías de sinceridad y, por consiguiente, de rigor.

Indicadores confirmados

Aunque cualquier experiencia, programa, institución o actividad de carácter educativo son diferentes a otros, existen hallazgos de la investigación que pueden servir de referentes para el contraste. Lo cual no significa que, si los resultados de una evaluación difieren de lo previamente descubierto, haya de considerarse ésta poco rigurosa. Puede servir para avivar la reflexión y revisar de forma concienzuda todo el proceso.

En relación con el soporte sobre el cuál establecer los juicios de valor, algunos autores recomiendan la utilización de indicadores de rendimiento (De Miguel, 1995). Entienden por tales, los datos empíricos –ya sean cuantitativos o cualitativos– recogidos de forma sistemática en relación con ciertas metas o procesos que nos permiten estimar la productividad y/o funcionalidad de un sistema. El procedimiento de los indicadores trata de dar sistematicidad a la recolección y credibilidad a los resultados. La formulación de los indicadores encierra una gran dificultad pero, sobre todo, el problema reside en el establecimiento de una doble correspondencia: la que se pretende conseguir entre los indicadores y los programas y la que su atribuye al programa concreto con los indicadores formulados.

Si se tiene en cuenta que cada programa es único e irrepetible, será difícil establecer algunos indicadores genéricos. La evaluación de los programas a través de la comprobación de los objetivos o la consecución de indicadores deja al margen preguntas de gran calado: ¿Y si los objetivos estaban mal planteados? ¿Si se consiguen de forma inmoral? ¿Si se logran de manera excesivamente costosa? ¿Si se impide que los protagonistas pro-

pongan sus propios objetivos? ¿Por qué no se han conseguido los objetivos? ¿Podrían haberse planteado otros mejores (más razonables, más justos)? ¿Si se han producido efectos secundarios negativos? ¿Para qué sirven esos objetivos? ¿Han existido medios suficientes para alcanzarlos? Otros problemas intrínsecos a la utilización de indicadores radican en los criterios de selección de los mismos, en su aplicación y en la interpretación de la información que aportan.

El control democrático de las evaluaciones

Cuando la evaluación se convierte en un juego que comparten sigilosamente los patrocinadores y los evaluadores, existen pocas garantías de credibilidad. Por eso resulta imprescindible que los informes se hagan públicos. Son los ciudadanos los que, en definitiva, tienen el control del proceso. Conocen no sólo los contenidos de los informes sino las características que ha tenido todo el proceso.

La devolución de informes a los protagonistas, la negociación de los mismos y la difusión del texto final a los ciudadanos es una parte fundamental del control democrático de la evaluación. Esta seguridad salvaguarda *a priori* de muchos abusos y, *a posteriori*, es un excelente medio para conocer y mejorar la práctica evaluadora. Cuando los evaluadores utilizan un lenguaje indescifrable para los evaluados y para las diversas audiencias interesadas por la educación, están robando el conocimiento a los ciudadanos. La evaluación se convierte así en un mecanismo de poder que utilizan a su antojo los que pueden y los que saben.

El control democrático de la evaluación no sólo se centra en el análisis de una experiencia aislada sino en el nivel macro de la cuestión. Los aspectos centrales sobre los que plantea el control podrían sintetizarse en las siguientes: ¿qué evaluaciones se hacen?, ¿quién decide que se realicen?, ¿por qué en ese momento preciso?, ¿qué finalidades las inspiran?, ¿qué principios éticos las guían?, ¿qué rigor tienen?, ¿a quién se encargan? ¿cómo y con quién se negocian?, ¿para qué se utilizan?, ¿quién las conoce?, ¿qué efectos producen?, ¿cuánto cuestan?, ¿quién tiene el control?

Mientras más amplia sea la extensión de los ciudadanos, más democrática será la metaevaluación. McDonald, en 1997, preguntaba en una sesión de trabajo celebrada en la Universidad de Málaga: “Quién decide, conoce, opina y utiliza la evaluación? La respuesta a esta cuestión se aproxima a las exigencias democráticas cuando es del siguiente tipo: todos”. En la

medida en que algunos privilegiados (sean éstos patrocinadores, evaluadores o informantes) tengan especial acceso a los datos o capacidad para hacer con ellos lo que deseen, se tratará de una evaluación autocrática. La participación de los ciudadanos en las cuestiones relativas a la educación (en todas sus fases: planificación, acción, evaluación, etcétera) es una garantía de su buen uso.

Por eso resulta importante que los evaluadores se comprometan con los patrocinadores a entregar informes a los evaluados y a otras audiencias. De no hacerlo así, el evaluador devolverá los informes al patrocinador que hará con esa información lo que considere oportuno.

La mejora

¿Para qué se realiza la evaluación? Considero que la respuesta a esta pregunta ha de situarse en la producción de conocimiento relevante que permita mejorar la bondad del programa. No se evalúa por el mero hecho de evaluar. De ser así, estaríamos ante un fenómeno de inequívoca inmoralidad. ¿Por qué desperdiciar esfuerzos, tiempo y dinero en una tarea estéril? Lo curioso es que siendo tan importante la finalidad, pocas veces se comprueba si realmente la evaluación ha facilitado y conseguido la producción de cambios importantes. ¿Quién ha de impulsar estos cambios? En primer lugar, los protagonistas de la acción que, al conocer lo que sucede, pueden efectuar cambios en lo que están haciendo. También ha de trabajar en la producción de cambios el patrocinador de la evaluación. Algunos pueden referirse directamente a su modo de actuar, otros se conseguirán con las ayudas que deben prestar a los protagonistas.

Este carácter ascendente del cambio rompe la línea jerárquica de la evaluación, que fácilmente podría convertirse en un ejercicio de puro control. Si se promueve y se descubre a través de ella que existen deficiencias en la forma de administrar las escuelas, no sería bueno olvidarse de esas exigencias y subrayar solamente las de los administrados. Una evaluación de este tipo se convertiría en un instrumento de dominio, en un mecanismo de poder que, en buena lógica, rechazarían los protagonistas de la acción. La evaluación, así planteada, interesaría a quienes gobiernan, pero sería vista como una amenaza por los profesionales.

No es fácil conocer qué cambios se producen a través de la evaluación, pero es imprescindible preocuparse por una cuestión tan básica. Los cam-

bios pueden plantearse en diversos niveles: en las prácticas profesionales, en las actitudes, en las concepciones y en las situaciones y las estructuras que las albergan. Hay algunos que pueden exigirse de inmediato y otros que necesitan un tiempo considerable. No se producen de manera automática sino que requieren un clima y unos medios. La forma en que se negocian, se realizan y se trabajan los informes de evaluación facilitará o dificultará el cambio deseado.

El poder de la evaluación radica en último término “en la movilización de las energías individuales” (Patton, 1978). No se evalúa por el mero hecho de evaluar. La pretensión fundamental es la transformación de las prácticas. Una de las preguntas que puede plantearse la metaevaluación se centra, precisamente, en la potencialidad que ha tenido para promover cambios. Los cambios no dependen exclusivamente de los evaluados. Algunos están supeditados a decisiones de los políticos y de los administradores. Sería interesante realizar este tipo de metaevaluación tiempo después de haber finalizado la evaluación. ¿Qué efectos ha producido?, ¿qué consecuencias ha tenido?, ¿qué cambios ha promovido?, ¿qué rastro ha dejado?

Una parte de estas preguntas se puede centrar en los efectos que ha tenido la evaluación para realizar nuevas evaluaciones. Algunas se realizan de tal manera que dejan el terreno completamente arrasado para que allí se vuelva a plantar una semilla semejante. Como he apuntado en otro lugar (Santos Guerra, 1994), el prisma del tiempo permite contemplar la realidad desde ángulos y en condiciones especiales. Se puede ver qué ha supuesto la prueba del tiempo, en el caso de que los cambios hayan sido fugaces o superficiales, se ha adquirido una nueva experiencia, se han podido generar condiciones de viabilidad, etcétera.

La transferibilidad y la dependencia

¿Sirven las conclusiones de una evaluación para otro contexto diferente? Una de las cuestiones fundamentales que se plantea la evaluación es la generalización de los resultados. Esta cuestión encierra, a nuestro juicio, tres matices diferentes:

- 1) La cuestión epistemológica que plantea la naturaleza científica de los hallazgos extraídos de un solo caso, que podría sintetizarse en la pregunta que sirve de título a un trabajo de Helen Simons: ¿existe ciencia de lo particular?

- 2) La cuestión ética de hacer público el conocimiento obtenido en la investigación, sobre todo si está financiada con fondos públicos: ¿es legítimo privatizar el conocimiento público?
- 3) La cuestión práctica sobre la transferibilidad de los resultados: ¿es posible aplicar a otros contextos los resultados obtenidos en la evaluación?

La evaluación naturalista no formula generalizaciones del tipo de las que surgen de los estudios racionalistas, experimentales, cuantitativos, articulados sobre procesos estadísticos que concluyen en afirmaciones o negaciones de significatividad. En la evaluación cualitativa de programas no se acaban formulando conclusiones “válidas”, con un determinado nivel de probabilidad, para todos los programas similares. Ninguna de las situaciones es exactamente igual a la otra. Las experiencias sociales no son replicables.

Transferibilidad

Cada contexto tiene su carácter concreto e irrepetible y de su estudio no se extrae la comprobación de hipótesis de carácter general. La experiencia es irrepetible en sí misma y es difícilmente aplicable en su integridad, ya que otro programa se moverá en otras coordenadas y en otro ambiente distinto. A los investigadores cualitativos les preocupa la validez, los cuantitativos hacen hincapié en la confiabilidad y la reproducibilidad de la investigación. No es posible repetir una investigación cualitativa, ya que las condiciones son cambiantes.

La *transferibilidad* es un proceso que nos permite aplicar los hallazgos conseguidos a otra situación diferente/similar. Cuando nos enfrentamos a otra situación, en parte distinta (ya que no es la misma) y en parte semejante (ya que tiene elementos comunes), nos preguntamos: ¿hasta qué punto son válidos en esta nueva situación los resultados obtenidos en la primera?, ¿son aplicables a este contexto aquellas conclusiones?, ¿se encuentran aquí los mismos procesos?

El primer requisito para alcanzar la validez externa es que exista un nivel suficiente de validez interna. Sin ésta última, no es posible hablar con un mínimo de rigor de la primera. Es decir, que si el trabajo no tiene consistencia interna, no será posible aplicar sus hallazgos a otra realidad, sea o no similar.

Cerraremos este apartado sobre la metaevaluación recordando algo que Glaser y Strauss (1967) subrayan al hablar de la credibilidad de los estudios cualitativos: son los lectores los que, en última instancia, han de juzgar sobre ella. Sería paradójico afirmar que la credibilidad de los estudios cualitativos puede fundamentarse de manera cuantitativa.

La dependencia

De acuerdo con Guba, la dependencia es la estabilidad de los datos obtenidos en la investigación. La estabilidad está amenazada por la naturaleza cambiante de los fenómenos que se estudian en el campo de la educación y de los mismos cambios que sufren los conocimientos y estados de ánimo del investigador, que es el principal instrumento de la exploración y del análisis. La dependencia exige la “rastreadibilidad”, es decir la posibilidad de explicar cómo se han producido los cambios, de justificar por qué se han modificado las decisiones y por qué se ha variado la utilización de los instrumentos que se había previsto o que se estaba realizando.

La ética en la evaluación

La metaevaluación no es un proceso exclusivamente técnico. Como la misma evaluación, tiene una dimensión sustancialmente ética, por eso, no se centra exclusivamente en la búsqueda de criterios de carácter metodológico o algorítmico sino que se ha de centrar en su vertiente moral. Es necesario mantener la vigilancia sobre esta cuestión trascendental porque no se trata de evaluar mucho, ni siquiera de evaluar de forma técnicamente perfecta sino de saber al servicio de quién y de qué valores se pone la evaluación. El quiebre de la ética puede darse en el fenómeno mismo de la evaluación, en su globalidad o en algunos de sus partes específicas. Se puede poner en marcha una evaluación al servicio del poder, del dinero o del prestigio de los patrocinadores, por ejemplo, y se puede obtener información de forma fraudulenta realizando entrevistas que están falsamente planteadas.

La ética no ha de presidir la forma de llevar a cabo el proceso sino que se preocupa por el respeto a los valores en el contenido de la evaluación. No sólo de los valores que existen en el programa o la institución que se evalúa, sino de los valores de una sociedad democrática. La autoridad legítima sobre el proceso de evaluación proviene del modelo normativo de la organización de la enseñanza en la sociedad. El modelo condiciona la

institucionalización de la enseñanza y promueve (o admite) uno de evaluación coherente, que no la contradiga. Por eso es necesaria la reflexión ética, ya que la evaluación puede convertirse en un elemento de transformación y de mejora del modelo hegemónico. La evaluación ha de poner de manifiesto las limitaciones éticas de los modelos imperantes para hacerlo avanzar hacia otras fórmulas más positivas.

La educación no es un fenómeno aséptico. La forma de organizar las escuelas tiene mucho que ver con la justicia social. Tampoco lo es la evaluación de los procesos educativos, sea cualquiera el ámbito y la temporalidad de los mismos. La ética que se exige a la evaluación ha de presidir también la metaevaluación. ¿Hasta dónde ha de llegar esta exigencia?, ¿quién evalúa a los metaevaluadores? Si se establecieran comisiones de garantía de la ética, ¿quién ejercería el control democrático sobre ellas?

No siempre se plantea la evaluación siguiendo estos criterios. Cuando se pretende medir los resultados de las experiencias y de los centros escolares a través de pruebas estandarizadas con el fin de hacer una clasificación según la posición en la escala y se conceden luego recursos siguiendo esa clasificación, se está utilizando la evaluación al servicio de la injusticia y de la desigualdad. Por muy rigurosas y precisas que sean las pruebas, no se evita el desajuste moral. Se está dando a los que más tienen (cultura, dinero, poder) todavía más a través de un instrumento que parece neutral y aséptico. Es una trampa terrible que algunos defienden de forma ingenua y muchos de manera interesada.

El aprendizaje en la evaluación

La metaevaluación implica que comprendamos cómo aprenden las personas a partir de la actividad de evaluación. Después de todo, sólo en el acto del aprendizaje se garantiza el valor de la evaluación. La evaluación permite construir un conocimiento que se cimenta en los datos recogidos durante la exploración y el trabajo que los evaluadores realizan con los mismos. Aunque la finalidad fundamental sea la mejora, lo cierto es que la evaluación aporta un caudal de nuevos conocimientos al acervo de saber sobre la educación.

Una fuente de información y de conocimiento es el proceso mismo de la evaluación. Se aprende, a través de la evaluación, cómo se debe planear, negociar, explorar, redactar informes, aplicar resultados, etcétera.

Resulta sorprendente la facilidad con que se repiten errores o dejan de utilizarse conocimientos ya contrastados sobre cómo se ha procedido al impulsar la política de evaluación o al desarrollar las evaluaciones concretas. ¿Quién puede realizar el aprendizaje? Los patrocinadores, los evaluados, los evaluadores, los teóricos y los ciudadanos en general.

¿Cómo puede realizarse este aprendizaje? En primer lugar, es necesaria una actitud de curiosidad y de incertidumbre que mantenga despierto el deseo de aprender. La certeza condena a la esterilidad. En segundo término, es preciso que se tenga acceso a la información, no solamente a los contenidos de los informes sino a los análisis del proceso. La *espiral de la comprensión* se va construyendo a medida que avanza el proceso y no sólo se apoya en la documentación escrita. Es más, no sólo parte de los aciertos y de los descubrimientos, sino que también puede surgir de los errores y de las limitaciones.

Referencias bibliográficas

- Angulo, F. (1993). "Metaevaluación del programa de educación de diabéticos en el Hospital Regional Carlos Haya de Málaga", en Soriguer Escofet, F. y otros, *La evaluación de programas de educación de pacientes. Una experiencia clínico pedagógica*, Sevilla: SAS.
- De Miguel, M. (1995). "Indicadores de rendimiento y evaluación de programas", en Pérez Juste, R. *et al.*, *Evaluación de programas y centros educativos*, Madrid: UNED.
- Glass, G. V. (1972). "Educational product evaluation: a prototype formal applied", en *Educational Researcher* 9, 1.
- Glaser, B.G. y Strauss, A. L. (1967). "Discovery of substantive theory: A basic strategy underlying qualitative research", en Filstead, W.I. (edit), *Qualitative methodology*, Chicago: Rand McNally College. Publis. Corp.
- Guba, E. (1981). "Criterios de credibilidad en la investigación naturalista", en Gimeno Sacristán, J. y Pérez Gómez, A., *La enseñanza: su teoría y su práctica*, Madrid: Akal.
- Hopkins, D. (1989). *Evaluation for School Development*, Londres: Open University Press.
- House, E. (1994). *Evaluación ética y poder*, Madrid: Morata.
- McCormick, R. y James, M. (1995). *Evaluación del currículum en los centros escolares*, Madrid: Morata.
- Patton, M. Q. (1978). *Utilisation-focused evaluation*, Beverly Hills: Sage Publications.
- Santos Guerra, M. A. (1993). *La evaluación, un proceso de diálogo, comprensión y mejora*, Archidona: Aljibe.
- Santos Guerra, M. A. (1994). "Quince años después. Evaluación de una experiencia bajo el prisma del tiempo", en Santos Guerra, M. A., *Entre bastidores. El lado oculto de la organización escolar*, Archidona: Aljibe.

- Santos Guerra, M. A. (1995). "Evaluación de los cambios organizativos en la reforma española de 1990", *Revista de Educación*, núm. 305.
- Santos Guerra, M. A. (1997). *La luz del prisma. Para comprender las organizaciones educativas*, Archidona: Aljibe.
- Santos Guerra, M. A. (1998). *Evaluar es comprender*, Buenos Aires: Magisterio del Río de la Plata.
- Santos Guerra, M. A. (2000). *La escuela que aprende*, Madrid: Morata.
- Santos Guerra, M. A. (2001). *Enseñar o el oficio de aprender*, Rosario: Homo Sapiens.
- Santos Guerra, M. A. (2003). *Una flecha en la diana. La evaluación como aprendizaje*, Madrid: Narcea.
- Santos Guerra, M. A. (coord.) (2003). *Trampas en educación. El discurso sobre la calidad*, Madrid: La Muralla.
- Scriven, M. (1976). "Evaluation bias and its control", en Glass, G. V., *Evaluation Studies Review Annual*, vol. 1, Beverly Hills: Sage Publications.
- Weston, A. (1994). *Las claves de la argumentación*, Barcelona: Ariel.