

# MONOGRÁFICO

## TENDENCIAS EN EVALUACIÓN

ERNEST R. HOUSE (\*)

Cuando, hace más de veinte años, comencé mi carrera como evaluador, recogí todos los artículos que pude encontrar sobre esta materia, los coloqué en una carpeta en un rincón de mi despacho y los leí en un mes. Desde entonces la evaluación de programas se ha ido transformando, de ser una actividad pequeña y marginal, llevada a cabo por académicos a tiempo parcial, a convertirse en una industria profesionalizada, con sus propias revistas, sus premios, convenciones, organizaciones y *standars*. El concepto de «evaluación», incluso, tiene su propia entrada en el *Dictionary of Modern Thought* (Bullock, Stallybrass y Trombley, 1988) (1).

Durante todo este tiempo, tanto las bases estructurales como los pilares conceptuales del campo de la evaluación han cambiado profundamente. Mientras que estructuralmente se ha ido integrando cada vez más en el funcionamiento organizativo de las instituciones, conceptualmente se ha desplazado desde la utilización de nociones monolíticas al pluralismo conceptual y a la utilización y la aceptación de métodos, criterios e intereses múltiples.

La evaluación se define normalmente como la determinación del valor de algo; en nuestro caso, de los programas, las políticas y el personal educativos, juzgados de acuerdo con criterios apropiados que han de ser, a su vez, explicados y justificados. En el mejor de los casos, la evaluación de programas sociales y educativos aspira a ser una institución para la democratización de las decisiones públicas, haciendo posible que los programas y las políticas estén más abiertos al escrutinio y a la deliberación públicos. En este sentido, la evaluación debería servir a los intereses no sólo de los patrocinadores, sino también de la sociedad, en

---

(\*) Universidad de Colorado en Boulder.

(1) El concepto de evaluación viene definido en dicho diccionario por Barry MacDonald, otro evaluador, en los siguientes términos: «En educación, el proceso de adquisición de información, normalmente para administradores y profesores, sobre los efectos y valores de las actividades educativas» (Citado por la edición inglesa; Bullock *et al.*, 1988, p. 291). (N. de T.)

general, y de los diferentes grupos que comprende. Pero por supuesto, la evaluación no siempre ha cumplido estas nobles aspiraciones.

## CAMBIOS ESTRUCTURALES

El más fuerte impulso para su desarrollo vino a través de la legislación promulgada por Lyndon Johnson para la «Gran Sociedad», que aunque no fuera capaz de cambiar la sociedad norteamericana en su totalidad, transformó, ciertamente, la investigación educativa. Debido a la insistencia del Senador Robert Kennedy, la «Ley sobre Educación Primaria y Secundaria» (*Elementary and Secondary Education Act*) (2) de 1965 aprobó la evaluación de los programas destinados a los niños con desventajas sociales, extendiéndose este mandato, con el tiempo, a todos los programas sociales e, incluso mucho más allá de las intenciones originales, a otros programas educativos (MacLaughlin, 1975). Antes de 1965, el término evaluación venía a significar el empleo de tests o evaluación del currículo; estando el trabajo concentrado en un puñado de especialistas en medición que se afanaban por mostrar la validez de los nuevos currícula propuestos. Después de 1965, el campo se parecía a los antiguos campos de oro de California, con un gran número de investigadores llegados de diferentes y dispersas disciplinas intentando ofrecer sus ideas sobre cómo deberían hacerse evaluaciones. Junto a especialistas en medición y en contenido, se podía encontrar psicólogos, sociólogos, administradores e incluso algunos filósofos. Por esto, y aunque la evaluación estuviera dominada inicialmente por el punto de vista psicológico de lo que era la investigación, la nueva mezcla de gentes, ideas y problemas estaba creciendo y madurando.

El campo se convirtió en multidisciplinar, con tradiciones separadas y múltiples opciones en desarrollo. Los evaluadores educativos suelen fechar históricamente sus comienzos en Ralph Tyler, los provenientes de la psicología en Donald Campbell y los de la sociología en Edward Suchman (3). Su experiencia previa con los tests y con el desarrollo del currículo supuso una ventaja para los primeros que les ha convertido en los teóricos más influyentes, una vez consolidada la disciplina. Este es, desde luego, un logro inusitado para los investigadores educativos, a quienes raramente se les concede un *status* intelectual semejante al de otros científicos sociales.

De esta extraña mezcla emergieron algunas escuelas de evaluación: Illinois, con su larga experiencia en medición y desarrollo curricular; Northwestern, con sus diseños experimentales «duros»; Western Michigan, con su hincapié en la toma de decisiones; UCLA, con el centro federal de investigación y desarrollo; y Stanford, con sus esfuerzos interdisciplinares. Los proyectos se multiplicaron,

---

(2) Véase nota 5.

(3) El autor se refiere a las siguientes obras: R. Tyler (1949) —aunque los primeros escritos sobre evaluación de este autor datan de 1934, una buena selección de sus trabajos se encuentra en Madaus y Stufflebeam (1989)—, D. Campbell (1969), Campbell y Stanley (1963) y E. Suchman (1967). (N. de T.)

los modelos de evaluación proliferaron y se erigieron nuevas organizaciones. El campo de la evaluación se desarrolló en este fermento intelectual y de contradicciones por más de una década, hasta que llegaron los recortes presupuestarios en los programas sociales. Los años de Reagan dieron pocas oportunidades a la evaluación, al menos a la evaluación externa y basada en contratos; entrando en una existencia más sobria y adormecida, muy distinta de la correspondiente a los anteriores días de la fiebre del oro.

No obstante, durante la apaciguada época de los ochenta, otros cambios se hicieron notar. La evaluación pareció convertirse en algo demasiado importante como para ser dejado en manos de los evaluadores. Las grandes burocracias desarrollaron sus propias agencias de evaluación. No sólo los distritos escolares y los departamentos estatales de educación, sino también organizaciones como el FBI, el Ministerio de Alimentación y Drogas y la Oficina General de Economía crearon sus propios equipos internos de evaluación. La aparición de estas unidades internas planteó de inmediato nuevos problemas de credibilidad y procedimiento, dado que gran parte de la teoría sobre la evaluación había sido construida y fundamentada en razón de relaciones contractuales entre evaluadores externos y organizaciones. ¿Frente a quién son responsables los evaluadores internos? ¿Qué grado de publicidad debe tener su trabajo? ¿Qué controles científicos son posibles? Éstas y otras cuestiones aparecieron a medida que la base estructural de la evaluación cambió.

Junto a ello, las nuevas reformas escolares iniciadas por las legislaturas estatales, durante los años de Reagan en la presidencia, marcaron un renovado énfasis en los tests estandarizados de rendimiento con el claro propósito de centralizar la exigencia de responsabilidades administrativas (4); por lo que, y de modo creciente, los tests se emplearon mucho más con propósitos de disciplina que de diagnóstico. La interpretación de los resultados de los tests se transformó claramente en un problema político, en el que los intereses ideológicos sustituyeron a las causas putativas del declive o el resurgir educativo. Incluso el currículum de las escuelas comenzó a reflejar el contenido que los tests medían. Los evaluadores educativos, que habían pensado que ya estaban lejos de considerar las puntuaciones de los tests como el único criterio de éxito, se encontraron, en comparación con tiempos pretéritos, con una desmesurada utilización de los mismos.

## CAMBIOS CONCEPTUALES

En parte como respuesta a las transformaciones estructurales operadas entre 1965 y 1990, la metodología, la filosofía y la política de la evaluación cambiaron también sustancialmente. La evaluación se desplazó de posiciones monolíticas a

---

(4) En este contexto, el significado de la expresión inglesa utilizada por el autor, *accountability*, está directamente conectado con una forma tecnocrática de valorar y burocrática de exigir el cumplimiento de obligaciones. (N. del T.)

concepciones pluralistas: una utilización de métodos, medicinas, criterios, perspectivas, audiencias e incluso intereses múltiples. Metodológicamente, transformó su énfasis inicial en los métodos cuantitativos, en los que los tests estandarizados de rendimiento empleados en los diseños experimentales con grupos de control elegidos al azar eran tenidos en la más alta consideración, a una actitud mucho más permisiva por la que los métodos de investigación cualitativos comenzaron a ser aceptados. Hasta el punto de que –ahora– el empleo simultáneo de métodos cualitativos y cuantitativos es algo rutinario en ese espíritu de ecumenismo metodológico en el que nos encontramos. Habiendo alcanzado la legitimidad necesaria en la comunidad científica, los evaluadores cualitativos han comenzado a disputar entre sí.

Filosóficamente, los evaluadores dejaron de creer que su disciplina era una disciplina neutral y asumieron que su práctica implicaba promover los valores e intereses de unos grupos sobre otros; aunque no acostumbran tener claro, en absoluto, qué hacer con semejante descubrimiento. Además, los evaluadores tuvieron que hacer frente a la disyuntiva, en apariencia conflictiva, entre ser científicos, por un lado, y útiles, por otro. Políticamente, por último, se movieron de una esfera en la que se concebían a sí mismos como técnicos expertos opuestos a los males de la política, a otra en la que ahora admiten, francamente, que la evaluación misma es una actividad con efectos e implicaciones políticos. Estos cambios conceptuales fueron estimulados por la rápida evolución del contexto social, en relación con el cambio del carácter en la sociedad norteamericana.

Tras la Segunda Guerra Mundial, los Estados Unidos alcanzaron una preeminencia en la política mundial raramente igualada en la historia de la humanidad. Agrupados en lo que Hodgson (1976) denominó «la ideología del consenso liberal», los americanos creyeron que su sistema de libre empresa tenía un claro potencial revolucionario para la justicia social; que la economía podría producir en cantidad suficiente para todos, de tal manera que los conflictos entre las clases, e incluso en las clases mismas, se verían eliminados; y que los problemas sociales podrían resolverse definitivamente de la misma forma que los problemas industriales. Una vez identificado el problema –se pensaba–, el Gobierno generaría (iluminado por la ciencia social) la solución pertinente, se aplicarían los recursos y los problemas se resolverían. La clave de este sistema revolucionario era el incansante crecimiento de la producción económica, que podría mantener suficientemente a todos.

En sus primeras formulaciones, a la evaluación se le asignó la tarea de descubrir qué programas educativos –tales como el *Elementary and Secondary Education Act (ESEA)*, el *Head Start* y el *Follow Through* (5)– funcionaban mejor. La mayoría

---

(5) El *Elementary and Secondary Education Act (ESEA)* (1965) y, concretamente, el *Title I* de esta ley fueron aprobados con el objetivo de hacer frente y paliar las necesidades educativas especiales de los alumnos en situaciones sociales y educativas de clara desventaja social, residentes en barrios marginales y pertenecientes a familias pobres. El *Head Start* se orientó a la educación preescolar y el *Follow Through (FT)* se constituyó como un «macroproyecto» formado en 1968 por doce proyectos (o mo-

de los evaluadores pensó que las ciencias sociales indicarían con claridad las causas de los problemas sociales y el tipo de intervenciones necesarias para superarlos; creyó que dichas intervenciones podrían ser aplicadas y evaluadas de tal manera que aportasen respuestas no ambiguas; y además esperaba que los resultados fueran aceptados con entusiasmo por patrocinadores, *managers* y políticos y que los problemas podrían resolverse definitivamente. Pero todas estas esperanzas se transformaron en frustraciones (Cook y Shadish, 1986).

La metodología que se consideraba como la mejor para realizar todas estas tareas fue la experimental, ampliamente alabada en trabajos seminales: «Este capítulo está comprometido con los experimentos como el único medio para zanjar las disputas sobre la práctica educativa y como la única vía de verificar las mejoras educativas...» (Campbell y Stanley, 1963, p. 2). Este enfoque se expresó durante muchos años en la política federal norteamericana y orientó el trabajo del evaluador hacia lo que Scriven (1983) denominó «las doctrinas de la gestión empresarial y del positivismo». Es decir, los análisis se condujeron hacia el beneficio de los gestores y directores de los programas y los evaluadores actuaron como científicos neutrales que depositaban su confianza en los métodos aceptados de las ciencias sociales como forma de protección frente a los sesgos. Las evaluaciones eran primariamente experimentales, cuasiexperimentales o empleaban el análisis de muestras, pero se utilizaban siempre mediciones de resultados cuantitativos con los que equipararse —por analogía— con las mediciones de desarrollo económico.

## CONJUNCIÓN METODOLÓGICA Y RUPTURA DEL CONSENSO

Los evaluadores encontraron pronto en sus estudios cuantitativos a gran escala muchos problemas que ni habían sido previstos, ni resultaban fácilmente superables. Los programas educativos variaban ampliamente de un lugar a otro, hasta el punto de que un programa como el *Follow Through* podía tener éxito en un lugar y fracasar en otro. Los modelos estadísticos (análisis de covarianza, etc.) «sobreajustaban» o «infraajustaban». Los participantes disputaban entre sí sobre los propósitos de diferentes programas. Los tests disponibles en la medición de resultados apropiados para un programa no lo eran para otro. A esto había que añadir el que muchos de los proyectos de reforma no tenían efectos consistentes. Las evaluaciones parecieron mucho más equívocas a la hora de aportar respuestas definitivas de lo que se había supuesto en un principio. Se cayó en la cuenta de que encontrar soluciones a los problemas sociales era mucho más difícil de lo que originalmente se pensaba.

---

delos, como se les denominaba) curriculares (desde el desarrollo del lenguaje hasta la enseñanza de las matemáticas). En 1969 se añadieron siete proyectos más. El macroproyecto *Follow Through* dio lugar en 1969 al *Head Start Planned Variation (HSPV)*, que mantenía la misma configuración experimental que el *FT*, con experimentos planificados de variación: «Programas diseñados para probar diferentes enfoques en la enseñanza infantil en diferentes lugares al mismo tiempo» (Rivlin y Timpane, 1975, p. 2). Para más información, además del texto citado pueden consultarse los trabajos de McLaughlin (1975), Rivlin (1971) y House *et al.* (1978). (N. del T.)

Es más, el consenso social que hasta entonces había existido en la sociedad norteamericana se fue desmembrando. Martin Luther King Junior, que comenzó su cruzada por los derechos civiles como una afirmación devota del sueño americano, terminó su vida en la desilusión y en su asesinato en 1968. Grupos disidentes se manifestaron en las calles y no solamente por los derechos civiles y contra la guerra del Vietnam, sino también por otras muchas causas, como por ejemplo, el feminismo, el poder negro, la liberación homosexual y la ecología. La economía nacional misma se tambaleó. Los ingresos medios familiares, tras haber alcanzado su cima en 1973, descendieron un 6 por 100 en 1984 y su distribución se hizo más desigual (Levy, 1987). Todo iba mal, incluida la «tarta» económica. Lo que comenzó siendo una era de consenso social se diluyó en una era de conflicto y diversidad.

La ruptura del consenso social se reflejó igualmente en los métodos de evaluación. Si diversos grupos esperaban diferentes cosas, entonces recoger las opiniones de la gente en y alrededor de los programas mismos parecía tener sentido. Los métodos cualitativos, útiles para esta labor, se pusieron de moda. Hasta entonces habían sido ampliamente utilizados en antropología y sociología, pero se pensaba que eran demasiado subjetivos para emplearlos en la evaluación de programas. Con el impulso de evaluadores como Robert Stake y Barry MacDonald, la metodología cualitativa desarrolló sus propios seguidores, una práctica y, eventualmente, su propia lógica de investigación. Al mismo tiempo, otros muchos evaluadores continuaron comprometidos plenamente con los métodos cuantitativos como aquellos métodos que había que elegir; con lo que las batallas entre defensores de una u otra metodología fueron muy reñidas durante los setenta (Boruch y Riecken, 1975; Eisner, 1979). Gradualmente, pero sólo tras dichas disputas, se alcanzó una cierta aproximación gracias a la cual la mayoría de los teóricos propugnaron métodos mixtos de investigación (Cronbach, 1982).

Pero aun así, la distinción entre métodos cualitativos y cuantitativos constituye, todavía, una marca importante de identidad entre los evaluadores. En un ejercicio de escala multidimensional, contrastando catorce teóricos de la evaluación, Williams (1989) encontró que la cuestión teórica más importante por la que se diferenciaban era su opción entre las metodologías cualitativa o cuantitativa. La segunda dimensión era la de responsabilidad *versus* asesoramiento. Esta dimensión indicaba si el evaluador debía juzgar programas y personal con el propósito de pedir responsabilidades administrativas o si debía adoptar un estilo más informativo. La tercera dimensión fue la de participación del cliente *versus* no participación, es decir, el grado y la forma en los que los clientes debían implicarse en la evaluación.

El debate metodológico, que ahora cumple veinte años, se extendió a otras cuestiones, además de la legitimidad, y a otras áreas de la investigación educativa. Pero actualmente, dicho debate se encuentra centrado en la cuestión de si los distintos enfoques metodológicos representan perspectivas diferentes del mundo o si son esencialmente compatibles (Guba y Lincoln, 1982; Smith, 1983). Muchos evaluadores aceptan el punto de vista según el cual, si no totalmente



conmensurables, ambos métodos son al menos complementarios; aunque nadie hasta ahora ha mostrado exactamente de qué manera métodos contrapuestos pueden conjugarse. Una cuestión relacionada con esto es la de qué *standards* resultan apropiados para los estudios cualitativos y si aquéllos pueden ser genéricos o si, por el contrario, han de ser derivados de las disciplinas de origen (Howe y Eisenhart, 1990).

Lo que subyace a estas disputas es el conflicto sobre la naturaleza misma de la causalidad, Cronbach (1982) ha puesto en cuestión la concepción *standard* de causalidad, aduciendo que los fenómenos sociales son menos predecibles de lo que la formulación *standard* suponía y que las nociones tradicionales de validez interna y externa tienen que ser redefinidas de tal manera que quienes utilicen los resultados de una evaluación puedan formular sus propios juicios y determinar, asimismo, la utilidad de la evaluación. Desde este punto de vista, la validez externa depende parcialmente de los juicios del usuario en el contexto y, por ello, resulta mucho más importante que la validez interna. Estas ideas todavía no han sido asimiladas y probablemente provoquen un debate mucho más extenso (6).

## LA UTILIZACIÓN DE LOS RESULTADOS

Otro área de interés durante los últimos diez años ha sido la de la utilización de los resultados de las evaluaciones. La idea según la cual los administradores y directores de los programas podrían aceptar fácilmente los resultados de las evaluaciones, ajustando o concluyendo, en su caso, los programas en concordancia con aquéllos, no ha sido, ciertamente, ratificada por el curso de los acontecimientos. Un intenso debate se ha generado sobre la cuestión de en qué medida han sido utilizados alguna vez los resultados de las evaluaciones. Weis (1988), por ejemplo, ha afirmado que el uso instrumental, es decir, la revisión de los programas como un resultado directo de la evaluación, raramente ocurre. Para Weis, el proceso de utilización de los resultados es un proceso de interpretación y esclarecimiento a través del cual dichos resultados se van incorporando gradualmente en las estructuras generales de referencia de los usuarios. Por el contrario, Patton (1988) ha insistido en la validez del uso instrumental, afirmando que los resultados serán usados si son presentados de forma apropiada.

Al parecer, algunas de las diferencias en cuanto al uso de las evaluaciones reside, probablemente, en el nivel jerárquico de decisión política en el que hayamos centrado nuestra atención. Mientras que los cambios en el nivel federal o estatal parecen producirse muy irregularmente y en relación con las fuerzas ideológicas y políticas predominantes, los cambios en los distritos escolares locales se generan de un modo mucho más instrumental (Smith, 1989). Una cuestión en la que los evaluadores sí están de acuerdo es la que plantea que si las evalua-

---

(6) Las últimas posiciones epistemológicas del autor se encuentran en House (1991). (N. del T.)

ciones pretenden ser útiles, deben ofrecer mejores descripciones de los programas y de sus contextos de actuación.

Otra idea emergente es la del uso erróneo de los resultados. Aunque, como acabo de decir, el debate se centró en un principio sobre el uso o el no uso de los mismos, se ha tenido que reconocer que estudios de evaluación pobremente concebidos habían tenido frecuentemente una amplia publicidad y que, por contra, los resultados de buenos estudios habían llegado a ser utilizados de forma impropia. Alkin (1990) ha propuesto una taxonomía de los usos erróneos de los resultados de evaluación en los que «uso» y «uso erróneo» se encuentran en los extremos opuestos de una escala en la que el «no uso» se sitúa en el centro. El hecho generalizado actualmente de que los resultados de tantos tests estandarizados de rendimiento se hagan públicos y que sus interpretaciones y la indicación de las cuestiones pertinentes resultantes se dejen en manos de los medios de comunicación y de los burócratas del Gobierno provoca que el uso erróneo sea, ciertamente, un acontecimiento particularmente destacado e importante.

De hecho, en los *standards* profesionales para la evaluación, desarrollados por un comité dirigido por Daniel Stufflebeam, se ha dedicado un considerable espacio a la cuestión del uso erróneo; pero el contexto en el que se presentan los resultados de las evaluaciones no suele orientarse por el empleo de dichos *standards*, incluso aunque hayan sido ampliamente aceptados por la comunidad de evaluadores (*Joint Committee*, 1981). Desde luego, lo que no está, bajo ningún concepto, claro es cómo atenuar o reducir este uso erróneo de los resultados de las evaluaciones.

## EL PAPEL DE LOS VALORES

Durante las dos últimas décadas, el creciente conflicto social ha colocado el problema de los valores en un lugar destacado. ¿De dónde proceden los valores en una evaluación? El acto de evaluación pública exige que sean establecidos algunos criterios de mérito y que estos criterios sean justificados. Tradicionalmente los objetivos propuestos en el programa han servido como fuente de criterios; centrándose la evaluación en el examen de si el programa había logrado, o no, dichos objetivos. Es más, aceptando que el patrocinador del programa fuera el cliente por excelencia de la evaluación, el evaluador podía actuar en razón de lo que era importante para aquél.

Sin embargo, muchos teóricos han cuestionado que la aceptación sin paliativos de los objetivos del programa sea la esencia de la evaluación. Scriven, en particular, ha estudiado la estructura de la lógica de la evaluación, señalando que la cuestión «¿es  $x$  bueno (malo, indiferente)?» y sus variantes (¿en qué medida  $x$  es bueno?, ¿es mejor que  $y$ ?) constituyen las preguntas evaluadoras prototípicas y que una respuesta a dichas preguntas necesitaría identificar y validar *standards* de mérito para  $x$ , así como descubrir las distintas actuaciones de  $x$  en las dimensiones que estuvieran relacionadas con el mérito (Scriven, 1980). De acuerdo con este razonamiento, los objetivos mismos del programa han de ser



también examinados. Por ejemplo, un evaluador responsable no aceptaría como criterio de evaluación de automóviles la suposición de la *General Motors* según la cual el mejor automóvil sería aquél que consiguiera mayores beneficios.

Esta lógica genérica todavía deja sin plantear la procedencia de los criterios particulares de mérito. Se puede afirmar, de modo general, que los criterios se derivan de lo que es apropiado para cuestiones de esa clase. Por ejemplo, no sería correcto afirmar que un currículum que pervirtiera la personalidad y retrasase el desarrollo intelectual constituiría un buen programa educativo, a pesar de que sus diseñadores quisieran provocar, justamente, ese efecto. Dada una entidad particular en un contexto particular, los criterios de mérito no pueden ser arbitrarios, sino que han de estar justificados. El fruto de este razonamiento es que los juicios valorativos no pueden ser tan arbitrarios como lo pueda ser un enunciado descriptivo del estilo de «un elefante es grande comparado con otros animales, pero pequeño en comparación con un moderno edificio de oficinas».

Por supuesto, el mundo social no es tan simple. Para entidades complejas como los programas educativos, existen múltiples y a menudo conflictivos criterios de mérito. Existen disputas en torno a los temas de aprendizaje inmediato o aprendizaje a largo plazo, conocimiento de hechos o pensamiento crítico, más historia o más matemáticas. Es más, la gente no siempre espera las mismas cosas de los programas públicos. Sus valores y, de hecho, sus intereses difieren. Un programa bueno para un grupo puede no serlo para otro. Incluso para el evaluador en acción, no existe otra opción que la de elegir entre los diversos criterios de mérito en controversia.

Muchos eligen las acostumbradas mediciones del rendimiento educativo, creyendo que éstas reflejan de la mejor manera posible los intereses generales. Otros sostienen que padres, estudiantes, administradores, profesores y ciudadanos deberían ser tenidos en cuenta a la hora de definir lo que es importante para la evaluación; solicitando sus opiniones, contando con ellos como parte de la audiencia o incluso implicándoles en la conducción del estudio mismo. Múltiples criterios, múltiples perspectivas, múltiples audiencias, múltiples intereses—concepciones pluralistas— reflejan el cambio del consenso al pluralismo que se ha dado en la sociedad en general.

Por supuesto, el tener múltiples criterios no resuelve el problema de cómo combinar estas multiplicidades en un juicio global. ¿Cómo «fusionar» los resultados de diferentes metodologías, por ejemplo? ¿Cómo combinar múltiples perspectivas? ¿Cómo compatibilizar los intereses contrapuestos de los grupos? Sobre estas cuestiones existe poco acuerdo. Algunos evaluadores creen que ellos mismos deberían juzgar; otros, que deberían solamente presentar los resultados y hacer que las diferentes audiencias formularan sus propios juicios. Unos proponen vías para resolver los conflictos de interés invocando técnicas como los análisis de costo-eficacia (Levin, 1983) y otros apelan, sin embargo, a procedimientos como el acuerdo negociado o el empleo explícito de teorías de la justicia (House, 1980).

## EL PAPEL DE LA POLÍTICA

Tras estos años, la evaluación ha sido aceptada como una actitud política; es decir, una actividad que se ve afectada por fuerzas políticas y que tiene, a su vez, efectos políticos. Las dos cuestiones críticas que se plantean en este ámbito son las siguientes: a qué intereses sirve la evaluación y cómo son representados en ella.

Como acabo de señalar, en un principio se asumía que los intereses de todas las partes eran reflejados adecuadamente por las tradicionales medidas de resultados. Esta suposición fue cuestionada y se reconoció que diferentes grupos poseían diferentes intereses y que podían ser afectados de forma también diferente tanto por el programa como por la evaluación. En la literatura inglesa ha comenzado a utilizarse el término *stakeholders*, en el sentido de interesados o implicados (7) (aquéllos que tienen algún interés, que se juegan algo, en el programa analizado), de tal manera que representar las visiones y los intereses de los implicados en la misma evaluación es ya una práctica aceptada.

El concepto de evaluación por implicados o interesados (*stakeholders*) se basa en la teoría actualmente predominante de la democracia como equilibrio y pluralismo de élites; teoría que renuncia a formular cualquier juicio normativo y que mantiene que los actuales sistemas de partes en competencia y grupos de presión tienen la función democrática de equilibrar las distintas y cambiantes demandas políticas (MacPherson, 1987). Se cree que la mejor manera de ajustarse al contexto político en el que los evaluadores operan consiste en describir lo que otros valoran, porque la toma de decisiones depende de los valores mantenidos tanto por los políticos como por los implicados e interesados relevantes. Presumiblemente, estas partes usarán los resultados para tomar decisiones informadas. Se supone que ni los administradores políticos ni los evaluadores intervendrán para apoyar un interés particular determinado, sino sólo para aportar información que sea objetiva y neutral respecto a los intereses y valores en juego. El interés de los diferentes grupos se disolverá de algún modo en los valores de los que tienen que tomar decisiones y de todos los implicados.

Sin embargo, este enfoque de evaluación posee importantes limitaciones. Por ejemplo, en dos decisivas evaluaciones llevadas a cabo según este enfoque y financiadas por el Gobierno federal, la de *Cities in Schools* (8) y la del programa de Jesse Jackson *PUSH/Excel* (9), los evaluadores trabajaron justamente en contra de los

---

(7) *Stakeholder* es un término inglés, procedente del mundo del juego, que viene a significar «corredor de apuestas». Sin embargo, en el contexto de la literatura especializada en evaluación se usa más bien en el sentido de «apostador», aquél que se juega algo. El término *stake* se usa para referirse tanto a apuestas como, en un sentido más genérico, a intereses. (N. del T.)

(8) *Cities in schools* era un conjunto de proyectos y programas urbanos de asistencia y ayuda a jóvenes residentes en *ghettos*. El primero comenzó en la ciudad de New York y fue patrocinado políticamente por el matrimonio Carter. (N. del T.)

(9) La idea de *Push for Excellence* fue el fruto de una campaña lanzada por J. Jackson, a finales de los setenta, en las escuelas urbanas de secundaria, para ayudar a los jóvenes negros a apartarse de la droga y a motivarles en el estudio. (N. del T.)

intereses de los participantes y de los estudiantes beneficiarios de los programas; grupos a los que, se suponía, iban a servir, poniendo en cuestión, por ello, la justicia misma de estas evaluaciones (House, 1988; Stake, 1986). Además, los resultados de la evaluación del proyecto *PUSH/Excel* fueron utilizados no solamente para desacreditarlo, sino también para poner en entredicho la habilidad y la capacidad de Jessie Jackson para coordinar grandes empresas políticas, precisamente de cara a la, por entonces, próxima campaña presidencial. Según estas experiencias, los «implicados» no tienen igual poder de influencia y uso de la evaluación y no gozan tampoco de igual protección frente a ella.

Los problemas de encontrarse con intereses múltiples y de cómo deberían ser representados colocan la evaluación en el ámbito de la justicia social. Aunque se reconozca la existencia de diversos implicados en una evaluación con legítimos intereses y, a menudo, en colisión, la cuestión de cómo conjugarlos permanece sin ser resuelta todavía. La práctica de describirlos de forma neutral parece inadecuada y la coyuntura actual, por la que gran parte de la actividad evaluadora se desarrolla dentro de grandes instituciones (estando sujeta, por tanto, a la autoridad burocrática), añade, desde luego, nuevas presiones. La decisiva cuestión política se mantiene: ¿A los intereses de quiénes debe servir la evaluación?

## SUMARIO

La evaluación se desarrolló originalmente como una estrategia para encontrar grandes soluciones a los problemas sociales, pero ésta ha resultado ser una aventura decepcionante y aleccionadora. El consenso ideológico inicial asumió que la solución para los problemas educativos se encontraba en la utilización de la metodología apropiada, que por entonces significaba medición cuantitativa de resultados empleando diseños experimentales, cuasiexperimentales o análisis muestral. Se aceptaba que diversos grupos sociales podrían verse representados a través de la medición clásica de rendimientos y objetivos educativos. Se asumía igualmente que los directores de los programas y los administradores políticos aceptarían de modo entusiasta los resultados de las evaluaciones, para poder así revisar o concluir el programa. Ninguno de estos supuestos ha resultado ser correcto.

A medida que el amplio consenso social se fue fragmentando, los evaluadores se volvieron hacia los métodos cualitativos, para poder recoger los puntos de vista de los participantes y de los clientes, es decir, de los «implicados» en los programas. Gradualmente, los evaluadores reconocieron que había muchos intereses que tener en cuenta en una evaluación y que algunos de ellos podrían estar en abierto conflicto entre sí. Como resultado emergió una concepción pluralista de la evaluación de programas; una concepción en la que los diversos y múltiples métodos, mediciones, criterios, perspectivas, audiencias e intereses se fueran considerando. Conceptualmente, la evaluación se desplazó desde una posición monolítica a otra pluralista, con la que se quería reflejar el cambio, del consenso al pluralismo, acaecido en la sociedad norteamericana. De qué manera

poder sintetizar, resolver y conjugar todas estas multiplicidades constituye todavía un problema formidable tanto para los evaluadores como para la sociedad en general. La evaluación, que fue inventada para solucionar nuestros problemas sociales, se encuentra afectada hoy en día, precisamente, por los mismos problemas que tenía que resolver.

## BIBLIOGRAFÍA

- Alkin, M. C., *Debates on evaluation*. Beverly Hills, SAGE Publishing, 1990.
- Boruch, R. F. y Riecken, W. H. (Eds.), *Experimental testing of public policy*. Bopulder, CO, Westview Press, 1975.
- Bullock, A.; Stallybrass, O. y Trombley, S., *The harper dictionary of modern thought*. New York, Harper and Row, 1988 (edic. rev.).
- Campbell, D. T., «Reforms as experiments». *American Psychologist*, 19, 1969, pp. 409-429.
- Campbell, D. T. y Stanley, J. C., *Experimental and quasi-experimental designs for research*. Chicago, Rand McNally, 1963 (traducción española en Amorrortu, Buenos Aires).
- Cook, T. D. y Shadish, W. R. Jr., «Program evaluation: The worldly science». *American Review of Psychology*, 37, 1986, pp. 193-232.
- Cronbach, L. J., *Designing evaluations of educational and social programs*. San Francisco, Jossey-Bass, 1982.
- Eisner, E. W., *The Educational imagination*. New York, McMillan, 1979.
- Guba, E. G. y Lincoln, Y. S., *Effective evaluation*. San Francisco, Jossey-Bass, 1982.
- Hodgson, G., *American in our time*. New York, Vintage, 1976.
- House, E. R., *Evaluating with validity*. Beverly Hills, SAGE Publishing, 1980.
- *Jesse Jackson and the politics of charisma: The rise and fall of the PUSH/Excel Program*. Boulder, CO, Westview Press, 1988.
- «Realism in research». *Educational Researcher*, 20 (8), 1991, pp. 2-25.
- House, E. R. et al., «No simple answer: Critique of the Follow Through Evaluation». *Harvard Educational Review*, 48 (2), 1978, pp. 128-192.
- Howe, K. y Eisenhart, M., «Standards for qualitative (and quantitative) research: A prolegomenon». *Educational Researcher*, 19 (4), 1990, pp. 2-9.
- Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects and materials*. New York, McGraw Hill, 1981.
- Levin, H. M., *Cost-Effectiveness: A primer*. Beverly Hills, Sage Publications, 1983.

- Levy, F., *Dollars and dreams: The changing american income distribution*. New York, Russell Sage Foundation, 1987.
- Macpherson, C. B., *The rise and fall of economic justice*. Oxford, Oxford University Press, 1987.
- McLaughlin, M. W., *Evaluation and reform: The elementary and secondary act of 1965/Title I*. Cambridge, MA, Ballinger, 1965.
- Madaus, G. F. y Stufflebeam, D. L. (Comps.), *Educational evaluation: Classic works of Ralph W. Tyler*. Boston, Kluwer Academic Publishers, 1989.
- Patton, M. Q., «The evaluator's responsibility for utilization». *Evaluation Practice*, 9 (2), 1988, pp. 5-24.
- Rivlin, A. M., *Systematic thinking for social action*. Washington, DC, Brookings Institution, 1971.
- Rivlin, A. M. y Timpane, P. M. (Comps.), *Planned variation in education: Should we give up or try harder?* Washington, DC, The Brookings Institution, 1975.
- Suchman, E., *Evaluative research. Principles and practice in public service and social action programs*. New York, Russel Sage Foundation, 1967.
- Scriven, M. S., *The logic of evaluation*. Inverness, CA, Edgepress, 1980.
- Scriven, M. A., «Evaluación ideologías», en G. F. Madaus, M. Scriven y D. L. Stufflebeam (Comps.), *Evaluation models: Viewpoints on educational and human services evaluation*. Boston, MA, Kluwer-Nijhoff, 1983, pp. 229-260.
- Smith, J. K., «Quantitative versus interpretive: The problem of conducting social inquiry», en E. R. House (Comp.), *Philosophy of evaluation*, San Francisco, Jossey-Bass, 1983, pp. 27-51.
- Smith, N. L., «The Weiss-Patton debate: Illumination of the fundamental concerns». *Evaluation Practice*, 10 (1), 1989, pp. 5-13.
- Stake, R. E., *Quieting reform: Social science and social action in an urban youth program*. Urbana, IL, University of Illinois Press, 1986.
- Tyler, R. W. (1949), *Principios básicos del currículum*. Buenos Aires, Troquel, 1973.
- Weiss, C. H., «Evaluation for decisions: Is anybody there? Does anybody care?». *Evaluation Practice*, 9 (1), 1988, pp. 5-19.
- Williams, J. E., *A numerical taxonomy of evaluation theory and practice*. Los Angeles, University of California Graduate School of Education, 1989 (mimeo).