



STVDIVM  
GENERALE  
CAESARAV-  
GVSTANAE  
CIVITATIS

COLECCIÓN PARANINFO  
SAN BRAULIO 2011



Universidad Zaragoza

LA CONSTRUCCIÓN  
DE LA INVESTIGACIÓN EVALUATIVA.  
EL APORTE DESDE LA EDUCACIÓN

Tomás Escudero Escorza



STVDIVM  
GENERALE  
CAESARAV-  
GVSTANAE  
CIVITATIS



Prensas Universitarias  
Universidad Zaragoza





LA CONSTRUCCIÓN  
DE LA INVESTIGACIÓN EVALUATIVA.  
EL APORTE DESDE LA EDUCACIÓN



LA CONSTRUCCIÓN  
DE LA INVESTIGACIÓN EVALUATIVA.  
EL APORTE DESDE LA EDUCACIÓN

Tomás Escudero Escorza

Prensas Universitarias de Zaragoza

- © Tomás Escudero Escorza
- © De la presente edición, Prensas Universitarias de Zaragoza  
1.ª edición, 2011

Prensas Universitarias de Zaragoza  
Edificio de Ciencias Geológicas  
c/ Pedro Cerbuna, 12 • 50009 Zaragoza, España  
Tel.: 976 761 330. Fax: 976 761 063  
puz@unizar.es <http://puz.unizar.es>

Impreso en España  
Imprime: INO Reproducciones  
ISBN: 978-84-15031-89-5  
Depósito legal: Z-???-2011

*A mis maestros*



## PREÁMBULO

Agradezco vivamente la oportunidad que me ofrecen el Rector Magnífico, profesor Manuel J. López Pérez, y su Consejo de Dirección de intervenir en el solemne acto académico del curso 2010-11, con motivo de la celebración de la festividad de San Braulio, patrono de nuestra Universidad de Zaragoza.

Intentando responder lo más adecuadamente posible a este honroso encargo, he seleccionado un tema de creciente relevancia social y académica y con gran incidencia en el quehacer de los investigadores, pero que por sus específicas características y, sobre todo, por sus consecuencias sigue siendo, con frecuencia, motivo de controversia y de debate científico.

La investigación evaluativa, objeto de mi análisis, se ha venido construyendo y formalizando como disciplina durante la segunda mitad del siglo XX, pero sigue siendo un proyecto en constante desarrollo, sobre todo en los ámbitos más aplicativos y de utilización de resultados. Actualmente, se trata de un enfoque investigador, en convivencia con otros más tradicionales, cada vez más habitual en proyectos de nuestros departamentos universitarios.

Aunque su origen, desarrollo y aplicación se ubican preferentemente en el terreno de los programas y políticas sociales, educación, salud, etc., y de las denominadas ciencias sociales, la investigación evaluativa incide hoy en día en la gran mayoría de las disciplinas académicas y campos del conocimiento.

La educación, una de las ciencias sociales por excelencia, por su gran apoyo en programas, instituciones, etc., ha sido la disciplina que, posiblemente, más ha aportado a la construcción, formalización teórica, desarrollo de modelos y metodologías de la actual investigación evaluativa. Sin embargo, es una disciplina que se mueve en el cruce de disciplinas; es una transdisciplina que se articula y engrana con cierta facilidad con los diversos ámbitos del conocimiento. Por este motivo, cada vez son más las titulaciones universitarias que incluyen a la investigación evaluativa, con una u otra denominación, en sus planes curriculares, bien como asignatura específica, bien como bloque de las asignaturas ligadas a metodología investigadora.

Los políticos deben tomar sus decisiones con los ojos abiertos; es tarea del evaluador iluminar la situación, pero no dictar la decisión.

(Tesis n.º 94, L. J. Cronbach y otros, 1980, *Toward reform of program evaluation*, Jossey-Bass, San Francisco)



# I INTRODUCCIÓN

La investigación evaluativa se ha convertido en las últimas décadas en una parcela muy relevante de actividad investigadora en todos los ámbitos científicos y de manera muy especial en las ciencias sociales, en general, y en educación, en particular. Cada vez son más los investigadores que trabajan en este campo, y muchos más los que incorporan en sus planteamientos sus principios, criterios y estrategias metodológicas.

La investigación evaluativa, también conocida en la literatura especializada como *evaluación*, se mantiene en constante evolución, pero tiene su periodo central de desarrollo en la segunda mitad del siglo XX, en buena medida como respuesta obligada a la necesidad de analizar, valorar y mejorar las políticas, acciones, planes, programas, instituciones y sistemas en los que se apoyan las sociedades modernas para desarrollarse y mejorar la calidad de vida de sus ciudadanos.

Durante el aludido siglo crece el impulso, desde los Gobiernos y otras organizaciones, de planes y programas ambiciosos de cambio y mejora social relacionados con la salud, la educación, la cultura, la protección social, la propia investigación científica, etc., que requieren ser analizados en profundidad para conocer su funcionamiento, su eficacia, su eficiencia, y para poder tomar decisiones

sobre su continuidad y su transformación. En este sentido, la investigación evaluativa se establece como un mecanismo estratégico de cambio social y una aliada obligada de los responsables políticos y sociales para optimizar sus acciones y sus decisiones. En este contexto, es lógico, tal como se hace hoy (McClintock, 2003; Calderon, 2004), definir al evaluador como un *investigador agente del cambio*.

Esta gran incidencia de los programas sociales en el desarrollo de la investigación evaluativa ha hecho que con frecuencia se siga hablando de *evaluación de programas* como denominación de este tipo de investigación. Sin embargo, el hecho de que el objeto de las evaluaciones sea más amplio que los programas, por ejemplo, políticas en general, instituciones, sistemas, recursos, personal, etc., ha servido para acuñar y consolidar la denominación más global y ajustada de *investigación evaluativa*.

Esta investigación adquiere una gran relevancia y desarrollo en ámbitos como la educación, la salud y el bienestar social, por la gran importancia que tienen para el funcionamiento y desarrollo de las sociedades modernas, pero debemos puntualizar que el alcance de este tipo de investigación es transversal e incide igualmente en el resto de disciplinas y ámbitos de relevancia social, como economía, cultura, ciencia, investigación, deporte, información, atención ciudadana, etc. Por este motivo, Scriven (1994) denomina a la investigación evaluativa *disciplina transversal* o *transdisciplina*.

En este *modelo transdisciplinar*, la investigación evaluativa tiene dos componentes, el conjunto de campos de aplicación de la evaluación y el contenido de la propia disciplina. En definitiva, la investigación evaluativa es una disciplina que incluye sus propios contenidos y los de otras muchas disciplinas; *su preocupación por el análisis y mejora se extiende a muchas disciplinas, es transdisciplinar*.

## EN LA ENCRUCIJADA DE LOS PARADIGMAS DE LA INVESTIGACIÓN

La investigación científica se basa en ciertos supuestos epistemológicos (Walker y Evers, 1988), que conducen a determinadas orientaciones y tradiciones investigadoras que hemos venido en denominar con Kuhn (1962) paradigmas, esto es, un *conjunto coherente de valores, postulados, fines, normas, lenguajes y formas de percibir, comprender y llevar a cabo la investigación*. En la práctica, muchas de estas orientaciones no son nítidas o puras, pues existen muchos cruces de posturas teóricas y metodológicas.

En el caso de las ciencias sociales en general, la literatura especializada ofrece un cierto acuerdo a la hora de detectar y destacar unas pocas grandes líneas paradigmáticas en la investigación (Popkewitz, 1984; Guba, 1990). Básicamente, se habla de tres grandes líneas: la *empírico-analítica*, de base positivista-racionalista; la *humanístico-interpretativa*, de base naturalista-fenomenológica; y la *crítica*, basada en la filosofía orientada al cambio social.

Algunos autores como Nisbet (1988) engloban los dos últimos paradigmas y añaden una línea paradigmática nueva orientada a la *toma de decisiones y el cambio*, que entronca de lleno con el fundamento de la investigación evaluativa, y que viene a coincidir con lo que Fishman

(1991) denomina *paradigma pragmático*, que más adelante analizamos, de naturaleza híbrida entre el *experimental* y el *hermenéutico*.

Estos paradigmas no siempre han tenido una convivencia pacífica; de hecho, los dos primeros surgen del conflicto de principios del siglo XX entre el positivismo clásico y el antipositivismo que aparece como reacción al primero. En síntesis, puede afirmarse que el debate epistemológico en ciencias sociales ha pasado por los tres momentos diferenciados siguientes (Smith y Heshusius, 1986):

- a) Mutuo rechazo entre paradigmas al principio.
- b) Se intentan analizar las diferencias entre los enfoques cuantitativos (positivistas) y cualitativos (humanistas), comparándolos bajo criterios racionalistas.
- c) Actitud de comprensión, compatibilidad y cooperación entre paradigmas. En la actualidad se defiende la complementariedad paradigmática, e incluso se busca la integración de paradigmas.

## 2.1. Enfoque positivista

El paradigma *empírico-analítico*, también denominado *positivista* y *cuantitativo*, sigue los métodos de las ciencias físico-naturales, consideradas como modélicas del conocimiento científico. Según Popkewitz (1984), los cuatro supuestos interrelacionados que dan forma a este paradigma son los siguientes:

- a) La teoría debe ser universal y no limitada a un contexto específico o circunstancial.
- b) La ciencia tiene un propósito analítico y sus afirmaciones deben ser independientes de los objetos y valores de las personas. La ciencia es neutral.

- c) El mundo social existe como un sistema de variables, que son diferentes y partes separables de un sistema interactivo.
- d) El conocimiento debe ser formalizado con variables seleccionadas de manera clara y precisa, conceptos operacionalizados, con unidades de análisis univariantes, para establecer variables, dependientes e independientes, sobre las que se estudia la dependencia mutua y los efectos de la manipulación de unas en las otras.

Apoyándose en la filosofía realista, defensora de que el mundo se puede captar tal como es, se mantiene la idea de que el conocimiento de la realidad solo es posible a través del método científico, se entiende que en el sentido positivista. Predomina el método deductivo y las técnicas cuantitativas. La indagación científica se produce formulando hipótesis, a partir de la teoría, cuyo valor se contrastará al comparar sus consecuencias deductivas con los resultados de las observaciones y de la experimentación controlada. Se buscan relaciones entre variables, a ser posible *relaciones causales*, que permitan predecir resultados manipulando las variables explicativas. Esta aproximación investigadora ha sido y es la más habitual, con aportaciones decisivas e insustituibles, lo que no resta relevancia a su incapacidad para resolver algunos problemas educativos y sociales, que demandan otro tipo de aproximaciones.

El positivismo radical, defensor a ultranza de cuestiones como las de la neutralidad y objetividad de la ciencia, resulta insostenible por su ingenuidad. Hoy en día, se habla generalmente de *post-positivismo* (Guba, 1990) para denominar a esta corriente paradigmática que mantiene sus preferencias por la metodología cuantitativa y la ciencia experimental pero con planteamientos mucho más

flexibles. Se acepta la variedad metodológica e incluso se recomienda la pluralidad y diversidad de fuentes y métodos como instrumento de validación científica.

## **2.2. Enfoque humanístico-interpretativo**

El segundo paradigma aparece en la literatura con diferentes denominaciones, apoyándose en algunos de sus principales elementos definitorios. Se habla de *paradigma humanístico-interpretativo*, constructivista, cualitativo, hermenéutico, simbólico, fenomenológico... Este paradigma intenta dar una respuesta a aquellos problemas inabordables con el anterior (Fetterman, 1984). Aunque se trata de un enfoque con tradición de siglos entre los pensadores sociales, su aplicación masiva como marco de investigación en estas ciencias es bastante reciente, en general como respuesta a las críticas al enfoque positivista durante varias décadas (Carr y Kemmis, 1983). A diferencia del empírico-analítico, más unitario en cuanto a fundamentación, este paradigma engloba una gran variedad de fuentes, posiciones y modos de acción, con bases en la fenomenología, el historicismo, el interaccionismo simbólico, etc. (Jacob, 1987; Cajide, 1992). Las fases cronológicas de introducción de estas tradiciones en la práctica habitual de los investigadores son analizadas por Bogdan y Biklen (1982), pasando por distintos niveles de antagonismo, encarnizado al principio, hasta posturas de complementariedad e integración (Walker y Evers, 1988), debido en gran medida a posiciones aperturistas y de apoyo hacia las metodologías cualitativas de algunos grandes científicos con origen claramente cuantitativista (Cronbach, 1957; Campbell y Stanley, 1963; Glass y otros, 1975). Bajo este enfoque, las investigaciones suelen rea-

lizarse en escenarios naturales y se abordan aspectos subjetivos de la conducta, con el apoyo de técnicas metodológicas como entrevistas en profundidad y la observación participante.

### **2.3. Enfoque crítico**

El tercer paradigma en la investigación es el que se apoya en la tradición filosófica de la *teoría crítica* (Habermas, 1971 y 1973), fundamentada en la apertura ideológica ante la interrogación científica, con orientación hacia el cambio social. Mantiene una flexibilidad total ante la diversidad metodológica, integrando métodos cualitativos y cuantitativos, análisis históricos y críticos. No se acepta la idea de una investigación neutral, no valorativa de los fenómenos que estudia, sino que obedece a un determinado interés al que sirve. Para Habermas, el saber humano surge por tres tipos de interés: el *técnico*, el *práctico* y el *emancipatorio*. El interés técnico persigue un saber instrumental que adopta la forma de explicaciones científicas y es característico de las ciencias empírico-analíticas o naturales. El interés práctico sirve para comprender y clarificar las condiciones para la comunicación y el diálogo significativos y generar un conocimiento interpretativo que guíe la práctica. Se trata de un saber de tipo hermenéutico e interpretativo. El tercer interés, el emancipador, pretende la autonomía racional y la capacitación para participar en la transformación social. Esto requiere la conexión contextualizada de hechos y valores. Se trata de un conocimiento autorreflexivo, propio de las ciencias críticas. Como vemos, estos tres intereses de la teoría crítica del conocimiento tienen una correspondencia clara con los tres paradigmas de investigación.

## 2.4. Controversia y convivencia entre paradigmas

La polémica sobre los paradigmas de investigación, que se ha centrado en gran medida en el debate entre lo cuantitativo y lo cualitativo, podría remontarse a tiempos antiguos, pero se empieza a formalizar en los términos actuales durante el siglo XIX, cuando se ponen de relieve las diferencias en el objeto de estudio entre las ciencias humanas y las naturales (Polkinghorne, 1983; Smith, 1983). Sin embargo, la controversia se agudizó en los años sesenta del siglo XX, cuando se vio la imposibilidad de verificar de manera rigurosa teorías utilizando únicamente datos cuantitativos (Alvira, 1983).

En el ámbito estrictamente educativo la convivencia entre los paradigmas positivista y humanista ha sido bastante buena, salvo en los años sesenta, en los que también se agudiza el debate (Husen, 1988). Es en esos momentos de crisis cuando van surgiendo nuevas propuestas paradigmáticas, bien alternativas del tipo de la basada en la teoría crítica (Guba y Lincoln, 1982), bien apoyándose en la complementariedad (Cook y Reichardt, 1986) y, en todo caso, ofreciendo posiciones diversas en cuanto a bases filosóficas, epistemológicas y metodológicas. En el interesante análisis de la conceptualización histórica de este debate que llevan a cabo Walker y Evers (1988) encuentran tres posturas básicas al respecto: *a)* la de *incompatibilidad* entre paradigmas, que compiten de manera irreconciliable; *b)* la de la *complementariedad* entre paradigmas, que, aunque de base epistemológica diferente, se apoyan y complementan en el proceso de investigación; y *c)* la de la *unidad* epistemológica en la ciencia, donde no se acepta la diversidad de paradigmas.

Los defensores de la *tesis de la oposición* o incompatibilidad de los paradigmas mantienen que son epistemoló-

gicamente diferentes y responden a concepciones distintas de la realidad social y también de las formas para llegar a conocerla (Smith y Heshusius, 1986). El dilema clave que plantean es que para el investigador cuantitativo existe una realidad externa, independiente de la mente, cuya observación suministra el fundamento epistemológico. Esto es negado por los cualitativistas, con lo que las diferencias están más en aspectos sustanciales sobre los criterios de verdad, objetividad, fiabilidad, validez, que en el hecho de utilizar más o menos información numérica y procedimientos estadísticos.

La *tesis de la diversidad complementaria* sostiene que los paradigmas son compatibles, no competitivos y que resultan complementarios, bien para problemas diferentes, bien para análisis desde diferentes perspectivas. Esta complementariedad fue defendida por Cronbach (1975), al indicar que los temas de investigación educativa son tan diversos que resulta imposible estudiarlos bien desde un único paradigma. Algo parecido defiende Soltis (1984), quien dice que la investigación educativa necesita a las tres tradiciones filosóficas habituales, a los estudios que buscan causalidad y relaciones, a los que buscan interpretar significados y a los normativos. Cook y Reichardt (1986) también apoyan la complementariedad paradigmática y metodológica y reclaman flexibilidad y diversidad en el uso de metodologías y técnicas de investigación.

La *tesis de la unidad* paradigmática es defendida por algunos autores (Keeves, 1988; Walker y Evers, 1988), aunque existe una gran diversidad de métodos que se complementan, permitiendo abordar los problemas desde múltiples perspectivas, dependiendo de la finalidad y circunstancias de la investigación. Tendrá que ser el investigador quien tome las decisiones más apropiadas.

das en este terreno (Shulman, 1988). En esta línea de pensamiento, pero yendo un poco más lejos, Tuthill y Ashton (1983) dicen que cada investigador debe desarrollar su propio paradigma, esto es, sus valores, supuestos, objetivos, normas, lenguaje, creencias y modos de percibir y comprender el proceso investigado. Howe (1985) apoya este tipo de argumentación cuando señala que una distinción epistemológica rígida entre los métodos cuantitativos y cualitativos se apoya en dogmas positivistas ya superados por el postpositivismo, como la elección forzada entre la investigación cualitativa apoyada en valores y la cuantitativa libre de valores. En un artículo posterior, Howe (1988) refuerza su anterior posición y proclama que la incompatibilidad entre paradigmas es fundamentalmente un artificio, porque, dice, lejos de ser incompatibles, los métodos cualitativos y cuantitativos están profundamente entrelazados. Quizás la posición más radical se encuentra en algunos autores como Maguire (1990), que aseguran que ninguna de las ciencias de la conducta ni ha tenido ni tiene un paradigma y que lo que se suele denominar paradigmas en estas ciencias son realmente *sectas metodológicas*.

Oberle (1991) se pregunta cuál ha sido el impacto de este debate de varios lustros en la investigación educativa. Responde, en primer lugar, que el positivismo radical está prácticamente agonizando, que se ha demostrado que la ciencia no está libre de valores, que las fuerzas sociales desempeñan un papel esencial en la manera como se observa y como se interpreta y que el progreso de la ciencia puede producirse de muchas maneras diferentes. Los filósofos de la ciencia han definido una nueva epistemología que reconoce el papel de las interacciones humanas en la empresa científica.

La mayoría de los implicados en el debate están a favor de la nueva epistemología, en la que cada paradigma tiene su contribución, con métodos cualitativos y cuantitativos en procesos de descubrimiento y de verificación. Algunos siguen manteniendo, desde posturas positivistas o interpretativistas radicales, la inconmensurabilidad de los paradigmas, pero estas posturas no saben presentar argumentos fuertes contra las posiciones epistemológicas intermedias, relativistas, aceptadas por muchos, a la vista de los argumentos y de las lagunas de las posturas radicales. En cuanto a la producción investigadora, sobre todo la que aparece en revistas especializadas, la investigación de orientación cuantitativa sigue siendo dominante. Eisner (1988) lo explica como una consecuencia lógica de lo costoso que resulta el proceso de socialización de las nuevas metodologías, venciendo una imagen social muy acentuada. Wachtel (1980) y Kupfersmid (1988) opinan que la razón está en gran medida en los sistemas de recompensas, publicaciones, que predominan en la comunidad científica. Pero resulta evidente que las metodologías cualitativas van ganando poco a poco crédito, incluso al nivel de publicaciones regulares en revistas especializadas. Sin embargo, estamos con Gage (1989) cuando dice que la batalla no ha terminado, porque esto no sucederá mientras existan investigadores que se apoyan más en dogmas que en concepciones reflexivas. La esperanza está, como también dice Gage, en que, si el debate sigue, ello sea en beneficio del avance de la investigación.

## **2.5. Características de los paradigmas**

La caracterización de los paradigmas se produce según las respuestas que dan sus defensores a cuestiones

básicas en la tres dimensiones siguientes: ontológica, epistemológica y metodológica (Guba, 1990; Oberle, 1991). La dimensión *ontológica* se pregunta por la naturaleza de la realidad, por la naturaleza de lo «cognoscible»; la *epistemológica* se pregunta por cómo se conoce, esto es, por la naturaleza de la relación entre el indagador y lo cognoscible; y, finalmente, la dimensión *metodológica* se pregunta por el proceso que debe seguir el investigador para descubrir lo cognoscible.

El axioma *ontológico* de los postpositivistas establece que la realidad social existe fuera de la mente de los sujetos, sin que pueda ser completamente aprehendida (realismo crítico). El objetivo de la ciencia postpositivista es exponer y articular las leyes naturales inmutables normalmente expresadas como generalizaciones y, por lo común, en forma de relaciones de causalidad. Para los constructivistas, la realidad es una construcción social múltiple; no hay una realidad única, tangible, fragmentable, sino que dicha realidad existe como una serie de construcciones holísticas, intra e interpersonalmente conflictivas y dialécticas en su naturaleza. Para el constructivismo la realidad es relativista y su interés es crear un conocimiento ideográfico, con hipótesis de trabajo, cuyo conocimiento está ligado al tiempo y al contexto. En la teoría crítica, al igual que en el postpositivismo, se acepta un realismo crítico.

La posición *epistemológica* del postpositivismo es dualista, en la medida que admite la separación entre el conocedor y lo conocido. Se trata de una relación neutral, libre de valores, en la que la objetividad se constituye como un ideal regulativo. Los constructivistas reemplazan esta objetividad por una subjetividad interactiva según la cual el investigador y lo investigado se fusionan en una única entidad (monismo). Finalmente, la posi-

ción crítica es subjetivista, puesto que la relación entre el sujeto y el objeto está marcada por la ideología. Se rechaza la neutralidad de la ciencia, se aceptan los valores explícitamente y la autorreflexión crítica en los procesos de conocimiento.

En cuanto a la *metodología*, el postpositivismo es intervencionista, se apoya en lo experimental, en lo manipulativo. Las corrientes interpretativas propugnan una metodología hermenéutico-dialéctica, focalizada en procesos sociales de construcción y elaboración. La metodología crítica es eminentemente participativa, pues las personas se comprometen e implican en el proceso de investigación. Se propicia una dialéctica intersubjetiva, con el fin de facilitar la concienciación y transformación de los participantes.

Dentro de cada paradigma, son preponderantes determinadas metodologías que, en definitiva, son las que plasman en la realidad de la investigación una u otra fundamentación. Las metodologías empírico-analíticas, racionalistas, cuantitativas, participan de los supuestos del postpositivismo y se suelen orientar al contraste de hipótesis, con el interés de elaborar un cuerpo de conocimiento que permita describir, explicar, predecir y, en lo posible, controlar los fenómenos. En general, se trata con fenómenos observables, susceptibles de medición y cuantificación, de control experimental y de análisis estadístico.

Las metodologías humanístico-interpretativas, histórico-hermenéuticas, se apoyan en una realidad más flexible y personal, creada por los propios sujetos, lo que obliga a tener en cuenta sus puntos de vista para captar significados e intenciones. Interesa la comprensión global de personas y situaciones, teniendo en cuenta la idiosincrasia de los fenómenos y su contextualización. Se uti-

lizan estrategias de observación interactivas, entrevistas etnográficas, análisis de relatos, etc., sin un filtro basado en definiciones conceptuales y operativas y sin escalas previamente estructuradas. Esto, que permite ciertos tipos de análisis imposibles en las metodologías empírico-analíticas, conlleva el problema de cierta debilidad metodológica en la información recogida.

La ciencia crítica en sus diversas formas y enfoques, con su orientación hacia el cambio social, mantiene una apertura total ante la diversidad metodológica, apoyándose en métodos cualitativos, cuantitativos, análisis históricos, críticos, y todo tipo de aproximaciones.

## EL PARADIGMA DE LA INVESTIGACIÓN EVALUATIVA

En el debate sobre los paradigmas de la investigación, aparece una corriente de analistas (Carr y Kemmis, 1983; Morin, 1985; Nisbet, 1988) que añade otro paradigma a los tres examinados anteriormente, orientado a introducir cambios en la práctica social y educativa. Estos autores ven al nuevo paradigma para el *cambio* como el único medio para resolver los problemas epistemológicos que tiene planteados la investigación desde una visión praxeológica caracterizada por la interacción entre acción y reflexión.

En este contexto y dentro de la corriente de pensamiento paradigmático de base crítica, aparece el paradigma denominado *pragmático*, de naturaleza híbrida entre los paradigmas experimental y hermenéutico, que formaliza Fishman (1991) y que, a nuestro entender, es el que mejor enmarca la realidad conceptual y metodológica de la investigación evaluativa. Se trata de un *enfoque reactivo* de la investigación que busca ofrecer soluciones a problemas concretos, desde una posición *pragmática* y *contextualista*, desde la que se defiende una noción situacional del desarrollo social, en vez de unos métodos estándar fijos como base y apoyo al desarrollo social emancipatorio.

Esta corriente filosófica, influenciada por Kant y Hegel desde el punto de vista idealístico, y por Darwin desde el materialístico (Bredo, 2006), adopta una visión evolutiva de la vida y de la indagación, y se rebela contra los fundamentos fijos y los esquemas deterministas y contra la noción de un fin último para la evolución natural y social. En este contexto de pensamiento, la división entre ciencia natural y social se suaviza, y se conciben los conceptos, las creencias y las normas como instrumentos vitales que deben ser reformados y adaptados con la experiencia (Escudero, 2009).

El nuevo paradigma, orientado hacia la resolución práctica de problemas, tiene, al igual que el hermenéutico, su apoyo en el construccionismo social, pero usa métodos cuantitativos como el experimental, así como indicadores de funcionamiento que apoyen el diseño de intervenciones eficaces. Siguiendo los principios del *pragmatismo* y el *contextualismo*, el paradigma pragmático se orienta hacia el logro de programas y proyectos que «funcionen bien» dentro de un contexto real particular. En el paradigma pragmático ya no es el propósito fundamental la descripción y construcción teóricas, sino la toma de decisiones para la gestión y la planificación, la solución de problemas y la construcción práctica de programas, dentro de la tradición investigadora aplicada (Fishman, 1991).

Apoyándonos en un análisis previo del propio Fishman (1988), en las tablas que se adjuntan ofrecemos una serie de similitudes y diferencias entre los tres paradigmas, experimental, hermenéutico y pragmático, en determinadas dimensiones y características. Conviene resaltar que en cualquiera de los criterios de comparación se trata de un continuo del que nosotros hemos seleccionado los extremos, y que existen muchas posiciones paradigmáticas intermedias para fundamentar la investigación.

SIMILITUDES Y DIFERENCIAS ENTRE LOS PARADIGMAS EXPERIMENTAL,  
HERMENÉUTICO Y PRAGMÁTICO

<b>Característica/ Dimensión</b>	<b>Paradigma experimental</b>	<b>Paradigma hermenéutico</b>	<b>Paradigma pragmático</b>
1. <i>Fundamentación epistemológica</i>	Positivismo lógico	Construccionismo social	Construccionismo social
2. <i>Modo básico de investigación</i>	Nomotético	Ideográfico	Ideográfico
3. <i>Finalidad principal de la investigación</i>	Derivación de leyes teóricas generales a través de la experimentación	Comprensión cualitativa de hechos y procesos en un contexto específico	Solución de problemas prácticos en contextos específicos
4. <i>Lugar principal para la investigación</i>	Situaciones específicamente creadas, como laboratorios, aulas, etc.	Situaciones naturales (reales)	Situaciones naturales (reales)
5. <i>Principales datos empleados</i>	Cuantitativos	Cualitativos	Cuantitativos y cualitativos
6. <i>Principal nivel de análisis</i>	Elemental, molecular	Holístico, molar	Molecular
7. <i>Principal fuente de conocimiento</i>	Observación	Intuición	Observación
8. <i>Visión de la predictibilidad y determinación de la conducta</i>	Énfasis en la conducta como determinada y predecible	Balance entre la conducta como predecible vs. no predecible y determinada vs. no determinada	Énfasis en la conducta como determinada y predecible
9. <i>Principal uso de la información empírica o experimental generada</i>	Desarrollo de la teoría científica	Enriquecimiento del conjunto acumulativo de comentarios acerca de la conducta y experiencia humana	Mejora de los programas sociales, asistiendo a los que toman las decisiones en dichos programas
10. <i>Principal modo de comunicación</i>	Publicaciones en revistas altamente especializadas	Publicaciones, revistas académicas y medios intelectuales	Informes donde apoyar la toma de decisiones acerca de programas y prácticas
11. <i>Necesidad de sistemas complejos de gestión de la información cuantitativa</i>	Relativa necesidad: la experimentación genera una cantidad relativamente pequeña de datos	Poca necesidad: los datos cuantitativos no son relevantes	Mucha necesidad: se trabaja con grandes cantidades de datos cuantitativos

CARACTERÍSTICAS COMPARTIDAS POR EL PARADIGMA PRAGMÁTICO  
CON OTROS PARADIGMAS

<b>Características compartidas por los paradigmas hermenéutico y pragmático</b>	<b>Características compartidas por los paradigmas experimental y pragmático</b>	<b>Características únicas del paradigma pragmático</b>
1. Compromiso con la epistemología del constructivismo social (vs. el positivismo lógico)	5. Los principales datos empleados son los cuantitativos (vs. cualitativos)	9. Énfasis en el conocimiento focalizado hacia la toma de decisiones
2. Énfasis de la investigación ideográfica (vs. la investigación nomotética)	6. El primer nivel de análisis es molecular (vs. molar u holístico)	10. Los informes se escriben frecuentemente para fundamentar la toma de decisiones en vez de solo para publicarse en revistas académicas
3. Énfasis en el conocimiento sobre un contexto específico (vs. las leyes generales)	7. La principal fuente de conocimiento es la observación (vs. el apoyo principal en la intuición)	11. Mucha necesidad de sistemas complejos de gestión de la información cuantitativa (vs. poca necesidad o no necesidad)
4. Principal ubicación de la investigación en el contexto natural (vs. laboratorio)	8. Énfasis en la conducta como predecible y determinada (vs. indeterminada e impredecible)	

LA INVESTIGACIÓN EVALUATIVA,  
DEFINICIÓN EN UN CONTEXTO  
DE CAMBIO SOCIAL

Cuando hablamos de investigación evaluativa, estamos tratando con un concepto complejo, con distintas orientaciones, que ha tenido y tiene un proceso constante de evolución. En el ámbito educativo, por ejemplo, es muy interesante observar este extraordinario proceso de evolución desde cuando la evaluación se identificaba con la medición de características personales hasta la investigación evaluativa actual (Escudero, 2003).

Pretender definir la investigación evaluativa con una simple frase, delimitándola de manera precisa de otros enfoques o ámbitos investigadores, resulta difícil porque existen muchas zonas de contacto que hacen borrosa la distinción. Así, por ejemplo, la investigación educativa aplicada y la investigación social tienen muchos elementos definitorios en común con la investigación evaluativa, aunque aparezcan matices diferenciadores. Solamente resaltando los elementos más específicos de la investigación evaluativa seremos capaces de establecer un *perfil definitorio válido y útil* para encuadrar el trabajo científico en este campo de desarrollo creciente.

Lo primero que queremos resaltar, y de ahí el título de este apartado, es que la investigación evaluativa se encuadra en un *contexto de cambio* y, más concretamente, en un

*contexto de cambio social*. Este hecho, elemento definitorio primordial, surge de la propia raíz paradigmática de la investigación evaluativa antes analizada (Mertens, 1999). Por ello, nos conviene recordar aquí los elementos definitorios del que anteriormente hemos definido como *paradigma pragmático*, híbrido entre el experimental y el hermenéutico (Fishman, 1991). Estos elementos son los siguientes:

- a) Fundamentación epistemológica en el *construccionismo social*.
- b) La *ideografía* como modo básico de investigación.
- c) La *solución de problemas* concretos como finalidad.
- d) Se investiga sobre todo en *situaciones naturales*.
- e) Se emplean tanto datos *cuantitativos como cualitativos*.
- f) El principal nivel de análisis es *molecular*.
- g) La *observación* es la principal fuente de conocimiento.
- h) Se enfatiza la *predictibilidad* de la conducta.
- i) Se busca la *mejora* de los *programas sociales*.
- j) Se informa a los responsables de *tomar decisiones* sobre programas y prácticas.
- k) Se trabaja con *grandes volúmenes de datos*.

#### **4.1. Algún apunte histórico**

Aunque este paradigma se conceptualiza como tal, al menos formalmente, más recientemente la investigación evaluativa comienza su andadura como disciplina académica a mediados del siglo xx. Kish (1987) señala que existían algunas actividades anteriores, pero data el nacimiento formal de la investigación evaluativa en 1955, en el *International Social Science Bulletin*. Años después empiezan a aparecer artículos y algunos manuales (Herzog, 1959;

Suchman, 1967; Wright, 1967; Weiss, 1972); mientras que hoy se trata de un campo académico muy pujante, con revistas y publicaciones múltiples, asociaciones científicas, presencia en los planes curriculares universitarios, etc.

Otros autores, como Gredler (1996), ubicarían el comienzo de la investigación evaluativa en 1932, con un estudio de seguimiento de ocho años en treinta escuelas (Eight-Year Study), dirigido por Ralph W. Tyler, pues lo consideran el primer estudio en el que se lleva a cabo una evaluación de programas. Otros se remontan al final del siglo XIX, en 1897, en el que aparece un trabajo de J. M. Rice al que señalan como la primera investigación evaluativa en educación (Mateo y otros, 1993). Se trataba de un análisis comparativo en escuelas americanas sobre el valor de la instrucción en el estudio de la ortografía, utilizando como criterio las puntuaciones obtenidas en los tests.

Hace más de cuatro décadas, Wright (1967) nos decía que existían dos tendencias claves en las actitudes modernas sobre evaluación. En primer lugar, la evaluación se contempla como un acompañante natural a los programas de acción social; y, en segundo lugar, existe un movimiento que demanda evidencias de éxito más sistemáticas, rigurosas y objetivas. *La aplicación de técnicas de las ciencias sociales para valorar los programas de acción social se ha venido en llamar investigación evaluativa.*

Wright nos sigue diciendo que *el sello de la investigación evaluativa moderna es la aproximación científica para evaluar el rendimiento de los programas.* En este sentido, la investigación evaluativa se equiparaba a otras formas de investigación social, pero la especificidad venía de los objetivos y de las condiciones de realización. Para entendernos más claramente, digamos que se consideraban programas sociales los de epidemiología y salud (Susser, 1975), pero

no estudios biológicos de laboratorio con animales, por ejemplo. *La investigación evaluativa, entonces, es un tipo especial de investigación social en relación con los programas sociales y con la toma de decisiones.*

Edwards y otros (1975) van más adelante, al decir que *la evaluación existe para facilitar la toma de decisiones inteligente*, de forma que, si no se cumple esto en el contraste entre programas que compiten, pierde su especificidad como investigación evaluativa y se convierte en simple investigación social.

Más recientemente, Kish (1987) analizó las principales notas distintivas de la investigación evaluativa, tomando como referencia de comparación la investigación social en general, y destacó las siguientes:

- a) La investigación evaluativa tiene que ver con los *programas de acción social* locales, nacionales e incluso internacionales.
- b) En la investigación evaluativa tiene lugar la *cooperación* entre una *agencia* responsable del programa de acción social y el *equipo de investigación* responsable de la evaluación. Pero se necesita cierta separación entre los investigadores y la agencia para facilitar la *objetividad* de la investigación y asegurar la *percepción pública* de la misma. A veces es necesario un patrocinador intermedio. Pero la colaboración entre todas las partes es vital. La agencia es responsable de los objetivos del estudio y el equipo de llevarlo a cabo bajo presupuestos científicos.
- c) La elección de *tratamientos y procedimientos* de observación está *condicionada* por las necesidades prácticas de evaluar la eficacia del programa estudiado. En este sentido, el evaluador no tiene tanta libertad práctica para diseñar su estudio como otros investigadores. Entre otros aspectos, el eva-

luador se ve obligado a trabajar en el contexto natural del programa.

- d) *La temporalización* suele ser un asunto delicado en la investigación evaluativa. En primer lugar, diremos que es difícil terminar la investigación y sugerir decisiones al respecto antes de que el programa se ponga en marcha, por mucho que esto sea deseable. Por otra parte, cambiar o desmantelar un programa en funcionamiento encontrará siempre obstáculos, incluso cuando falla, porque las críticas son una medicina amarga para los participantes. Además, cuando un programa afecta a una población muy amplia, el control empírico se complica.
- e) La evaluación existe para facilitar la toma de decisiones inteligente, pero estimar la *eficacia* de un programa es algo más que la medición de sus efectos. No basta con detectar efectos, hay que precisar su tamaño. A menudo los efectos de un programa son múltiples, incluso negativos, y estos deben ser confrontados con los positivos. También hay que buscar la detección de efectos no esperados (*serendipity*). Hay que analizar la eficacia en relación con los *costos*, *eficiencia*, y desde múltiples perspectivas.
- f) La difusión de la investigación evaluativa sigue unos cauces diferentes de los de la investigación convencional. A la agencia se le debe informar a plazo fijo, mientras que al gran público se le suele retardar o incluso esconder la información, muy a menudo reservada, en el caso de la investigación evaluativa.

Las características antes mencionadas de la investigación evaluativa, limitaciones, decisiones, rapidez, etc., a veces entran en conflicto con las normas de la investigación académica. Los equipos evaluadores necesitan

independencia y objetividad, real y en cuanto a imagen. Para lograr esto se han probado distintas formas de organización, como los equipos aislados de investigación dentro de agencias públicas, con independencia y reputación garantizadas (Converse, 1986), pero esto es un terreno en el que siguen existiendo puntos abiertos. A nosotros, por ejemplo, nos vienen a la mente los problemas, dudas y debates sobre las posibilidades de ubicación y de entronque institucional en el proceso de creación de las Agencias Estatal y Autonómicas de Evaluación en España.

#### **4.2. En interacción con el entorno sociopolítico**

La profesora Carol H. Weiss, uno de los máximos referentes en el campo de la investigación evaluativa, sobre todo en los primeros setenta (Weiss, 1970, 1972 y 1975), no solo se alinea con posiciones parecidas a las que acabamos de describir, sino que va un poco más allá, enfatizando el obligado *entorno político* de la investigación evaluativa. No basta hablar de entorno social, es necesario hablar de entorno sociopolítico, nos viene a decir. En primer lugar, porque la investigación evaluativa trata con objetos, programas, instituciones, etc., que son *criaturas políticas*, de decisión política. En segundo lugar, porque la evaluación se hace para alimentar la toma de decisiones, y, por lo tanto, sus *informes entran en la arena política*, compitiendo con otros elementos y factores intervinientes en el proceso político. Y, en tercer lugar, algo que según Weiss es menos reconocido, porque la evaluación tiene un sentido político en sí misma. Por su propia naturaleza, la evaluación toma posiciones políticas implícitas sobre la naturaleza de los programas, la legitimidad de sus objetivos, de sus estrategias, la utili-

dad de las estrategias de reforma propuestas, e incluso la influencia lógica del científico social en la política sobre los programas.

Este último argumento de Weiss entendemos que puede matizarse, pero creemos que ni el evaluador se mueve en un *contexto neutro*, ni el programa, sistema, institución u objeto que analiza son inmunes a presiones del contexto propio del mismo. En el centro de la escena está el apoyo, la oposición y el pacto político, el prestigio de legisladores, administradores y educadores y las expectativas de todos los implicados (Escudero, 1980).

La grandeza y el riesgo de la investigación evaluativa está, como señala House (1973), en que es un medio que, en una sociedad pluralista, utiliza un grupo para justificar su comportamiento ante otros. Reconociendo esa realidad, los investigadores tienen que extremar sus posiciones en defensa de su independencia y objetividad y contra la utilización tendenciosa de sus trabajos, empleando estrategias adecuadas para ello (Chelimsky, 2008).

Inciendo también en esta línea, Berk y Rossi (1990) señalan al menos tres conceptos claves en la investigación evaluativa: el *espacio político* que enmarca el dominio de las decisiones; los *implicados (stakeholders)*, responsables políticos, participantes, responsables administrativos, etc.; y la *eficacia de los programas*, con sus múltiples perspectivas de análisis, tanto internas como externas.

Mertens (1999) defiende arduamente el papel de la investigación evaluativa y de los evaluadores en la promoción de cambios sociales justos, a favor de los discriminados y de los más desfavorecidos. Defiende la *inclusión* como *stakeholders* de los tradicionalmente ignorados, pero nos dice que no basta con incluir su voces en la evaluación; se necesita un paso más. Sus voces proporcionan más credibilidad a la evaluación, pero la transformación

solo puede ocurrir realmente si su información es usada para fundamentar programas sociales que intenten atajar las desigualdades existentes.

Desde una posición global y comprensiva de este tipo de planteamientos, una posible definición actualizada de la investigación evaluativa, sería la siguiente (Escudero, 2006: 271): «*un tipo de investigación aplicada, que incide sobre objetos sociales, sistemas, planes, programas, participantes, instituciones, agentes, recursos, etc., que analiza y juzga su calidad estática y dinámica según criterios y estándares científicos rigurosos múltiples, externos e internos, con la obligación de sugerir acciones alternativas sobre los mismos para diferentes propósitos como planificación, mejora, certificación, acreditación, fiscalización, diagnóstico, reforma, penalización, incenti-vación, etc.*».

#### **4.3. Precisiones sobre la denominación**

Esta definición que acabamos de ofrecer de la investigación evaluativa es muy concordante con las múltiples que nos encontramos en la literatura especializada con esta misma denominación (McMillan y Schumacher, 1984; Arnal y otros, 1992) o con la denominación de *evaluación de programas* (Colás y Rebollo, 1993), limitando a los programas los citados objetos sociales. De la Orden (1985), por ejemplo, que señala que el término *investigación evaluativa* aparece por primera vez en el *Educational Index* en 1974, define la evaluación de programas como «el proceso sistemático de recogida y análisis de información fiable y válida para tomar decisiones sobre un programa educativo. Este proceso, en la medida en que es riguroso, controlado y sistemático, constituye un modo de investigación que hoy se conoce con el nombre de investigación evaluativa» (134).

No deseamos profundizar mucho más en este debate sobre la terminología, que en el fondo nos parece un poco equívoco puesto que, en la práctica, todos los investigadores evaluativos en ciencias sociales y educación trabajan con más objetos de evaluación que los programas (Nevo, 1989). Incluso desde una posición restrictiva, estos los debemos entender como una denominación genérica de multitud de objetos de evaluación posibles. Para nosotros, desde luego, la investigación evaluativa es mucho más que la evaluación de programas, aunque reconozcamos que esta se encuentra en el origen y en el núcleo de aquella. Hoy en día, quizás una de las posiciones más globales, al menos en el terreno educativo, está en el concepto de *evaluación institucional*, que, lógicamente, incluye la evaluación de su realidad, de sus necesidades, de su contexto, de su organización, de sus programas, de su personal, de sus recursos y de sus logros.

El que fue primer presidente de la American Evaluation Association y probablemente el principal responsable de nuestros grandes debates sobre *evaluación formativa vs. sumativa* (Scriven, 1967) nos ofrece un interesantísimo ensayo reconceptualizando lo que nosotros conocemos como investigación evaluativa (Scriven, 1994) y haciendo especial hincapié en el marco común de todo tipo de evaluaciones, por ejemplo entre la de personal y la institucional. Efectivamente, nos dice, existe una *disciplina general* que tiene varias áreas específicas desarrolladas autónomamente. De ellas, destacan algunas, como las que Scriven denomina las seis grandes (*Big Six*), esto es, la evaluación de programas, de personal, de rendimiento, de productos, de proyectos y de políticas. Ciertamente, resulta difícil definir mejor la evaluación institucional. Además, Scriven destaca otras dos grandes áreas de la evaluación, la que se produce dentro de cada

disciplina académica, con sus particularidades y especificidades, y la evaluación de las evaluaciones, esto es, la *metaevaluación*. Otras áreas citadas, además de estas ocho, son la evaluación curricular, la de la tecnología, la ética médica, la jurisprudencia, el control de calidad industrial y otras varias de menor alcance. Nevo (1989), en un ensayo anterior sobre la conceptualización de la evaluación, también señala la multiplicidad de objetos de estudio para la misma.

El que esta disciplina general no se haya asentado definitivamente hasta hace no muchos años, a pesar del largo uso de la denominación, se debe, según Scriven, al empleo restrictivo que se ha hecho de la misma. Así, por ejemplo, el término *evaluation* ha aparecido de manera genérica en muchos títulos de libros. Habría que haber supuesto que se referían a una disciplina general. «Patéticamente», al abrir los libros, durante seis décadas, estos se referían a evaluación (*assessment*) de los estudiantes, esto es, a una parte de un área aplicada en una disciplina académica (*evaluación del rendimiento educativo*). Más recientemente, el término genérico en el título de los libros se refería simplemente a la *evaluación de programas*, con escasas referencias a la evaluación de, por ejemplo, centros, profesores, administradores, recursos, etc. Con estas posiciones reduccionistas, sin el marco de una disciplina general que acoja todo tipo de investigaciones evaluativas, es lógico que hayan persistido planteamientos parciales, modelos incompletos y zonas de confusión abiertas.

Como respuesta a estos problemas, Scriven plantea un modelo que denomina *transdisciplinar*, del que hablamos más adelante. La evaluación, nos dice, pertenece a un reducido grupo de disciplinas que son *transdisciplinas*, que tienen entre una de sus principales funciones proporcionar instrumentos a otras disciplinas, aunque tam-

bién poseen un estatus identificativo propio. Además de la investigación evaluativa, pertenecen a este grupo disciplinas como la estadística, la medición, la teoría de la decisión y la lógica. Esto es completamente diferente de las *crossdisciplinas* o cruces interdisciplinares, que emergen o son creadas como híbridos o cruces de dos o más disciplinas concretas (Scriven, 1998).

Clarke y Dawson (1999) comienzan su texto sobre investigación evaluativa con una amplia reflexión sobre el concepto de evaluación y las distintas posiciones que se han ido sucediendo y se mantienen al respecto, y, que, en definitiva, configuran en la actualidad la disciplina. Se señala la tendencia creciente a utilizar como intercambiables los términos *evaluación* e *investigación evaluativa*, pero se reconoce la existencia de algunas posiciones que marcan algún punto de distinción entre ambos. También se reconoce que es dominante la posición de que la investigación evaluativa debería ser tratada como un área especializada de la investigación social aplicada, pero siguen existiendo otras concepciones, como la transdisciplinar de Scriven o la de House (1993: 77), que habla de una disciplina específica, aunque de perfiles algo difusos (una «casi-disciplina»).

Sintetizando de alguna manera este panorama definitorio y contextualizador de la investigación evaluativa, podemos decir lo siguiente:

- Se trata de un tipo de investigación aplicada que por su orientación, contenido y contexto de realización tiene unas características específicas que la identifican como disciplina (Escudero, 2005-2006).
- Es una investigación que incide en todo tipo de disciplinas y en las prácticas que resultan de ellas, moviéndose desde niveles muy prácticos hasta el nivel conceptual.

- Estudia objetos muy diversos y relevantes para el desarrollo de las políticas sociales. En educación, su campo natural de trabajo son los planes y programas, los centros y las instituciones, los estudiantes, los profesores, los responsables educativos y todo tipo de recursos.
- Los distintos campos de la investigación evaluativa, programas, centros, personal... tienen muchos niveles de interconexión y solapamiento.
- Se estudia y juzga la calidad de los objetos evaluados desde parámetros y criterios científicos variados y desde diversas perspectivas.
- Los potenciales propósitos de la evaluación son múltiples y variados.
- Se hace para que sea útil, para que sirva a los distintos implicados, con sugerencias alternativas de acción para la toma de decisiones.

Vemos que la investigación evaluativa se plantea siempre en un contexto real, al servicio de las políticas sociales, en definitiva, al servicio del cambio y del desarrollo social, y que, por lo tanto, es un instrumento intermedio, de ayuda a otras disciplinas y ámbitos como la educación, la sanidad, la cultura, el bienestar social, etc., pues se plantea directamente el objetivo de ofrecer las mejores opciones de acción entre todas las alternativas posibles. Se trata de ayudar a resolver problemas que aparecen en estos ámbitos. La investigación evaluativa, en suma, se mueve en un contexto de resolución de problemas. Y en este sentido, el evaluador se ve obligado a utilizar el conocimiento teórico en la resolución de problemas prácticos, desempeñando el importantísimo papel de facilitador de la integración de la investigación básica y la práctica (Urban y Trochim, 2009).

Es indudable que la educación, por su enorme relevancia social, ha sido y es uno de los principales objetos de interés para la investigación evaluativa, pero, al mismo tiempo, ha sido desde la educación desde donde, probablemente, más se ha impulsado el desarrollo de la propia investigación evaluativa, con una evolución constante y continuada a lo largo de muchas décadas (Escudero, 2003), pero con algunos momentos especialmente importantes que conviene resaltar.

### **5.1. La revolución tyleriana**

Nuestro primer referente debe ser Ralph W. Tyler, quien es tradicionalmente considerado como el padre de la moderna evaluación educativa, por ser el primero en dar una visión metódica de la misma, superando desde el conductismo, muy en boga en el momento, la mera evaluación psicológica. Entre 1932 y 1940, en su famoso *Eight-Year Study of Secondary Education* para la Progressive Education Association, publicado dos años después (Smith y Tyler, 1942), plantea la necesidad de una evaluación científica que sirva para perfeccionar la calidad de la educación. La obra de síntesis la publica unos años des-

pués (Tyler, 1950), exponiendo de manera clara su idea de «curriculum», e integrando en él su *método sistemático* de evaluación educativa, como *el proceso surgido para determinar en qué medida han sido alcanzados los objetivos previamente establecidos* (véase también Tyler, 1967 y 1969).

Esta evaluación ya no es una simple medición, porque supone un *juicio de valor* sobre la información recogida. Se alude, aunque sin desarrollar, a la toma de decisiones sobre los aciertos o fracasos de la programación, en función de los resultados de los alumnos, algo que retomarán otros grandes evaluadores como Cronbach y Stufflebeam unos años después.

## **5.2. La eficacia de los programas y la rendición de cuentas**

Los años sesenta traerán nuevos aires a la evaluación educativa, entre otras cosas porque se empezó a prestar atención, tras las indicaciones de Tyler, a la eficacia de los programas y al valor intrínseco de la evaluación para la mejora de la educación.

De la misma forma que la proliferación de programas sociales en la década anterior había impulsado la evaluación de programas en el área social, los años sesenta serán fructíferos en demanda de evaluación en el ámbito de la educación. Esta nueva dinámica en la que entra la evaluación hace que, aunque esta se centraba en los alumnos como sujeto que aprende, y el objeto de valoración era el rendimiento de los mismos, sus funciones, su enfoque y su última interpretación varíe según el tipo de decisión buscada.

Buena parte de este fuerte ímpetu evaluador se produjo en el sistema educativo americano con la aprobación de la Elementary and Secondary Act (ESEA) en

1965 (Berk, 1981). Con esta ley se puso en marcha el primer programa significativo para la organización de la educación en el ámbito federal de los Estados Unidos, y se estipuló que cada uno de los proyectos realizados con el apoyo económico federal debía ser anualmente evaluado, a fin de justificar subvenciones futuras.

A finales de los años sesenta, y como consecuencia, en parte, de la dura recesión económica, entra en escena un nuevo movimiento, la era de la *Accountability*, de la rendición de cuentas (Popham, 1980; Rutman y Mowbray, 1983), que se asocia fundamentalmente a la responsabilidad del personal docente en el logro de objetivos educativos establecidos.

Como consecuencia de esta nueva orientación, se amplió considerablemente el fenómeno de la evaluación educativa. El sujeto directo de la evaluación siguió siendo el alumno, pero también todos aquellos factores que confluyen en el proceso educativo (el programa educativo en un sentido amplio, profesor, medios, contenidos, experiencias de aprendizaje, organización, etc.), así como el propio producto educativo. En este contexto, se inicia un periodo de reflexión y de ensayos teóricos con ánimo de clarificar la multidimensionalidad del proceso evaluativo. Estas reflexiones teóricas enriquecerán decisivamente el ámbito conceptual y metodológico de la evaluación, lo que, unido a la tremenda expansión de la evaluación de programas ocurrida durante estos años, dará lugar al asentamiento de lo que se terminaría denominando *investigación evaluativa*.

### **5.3. La influencia de Cronbach y Scriven**

Como hitos de la época hay que destacar dos ensayos por su decisiva influencia: el artículo de Cronbach

(1963) «Course improvement through evaluation» y el de Scriven (1967) «The methodology of evaluation». La riqueza de ideas evaluativas expuestas en estos trabajos nos obliga a referirnos a ellos, aunque sea brevemente.

Del análisis de Cronbach del concepto, funciones y metodología de la evaluación, entresacamos las sugerencias siguientes:

- a) *Asociar el concepto de evaluación a la toma de decisiones.*  
Distingue el autor tres tipos de decisiones educativas a las cuales la evaluación sirve: 1) sobre el perfeccionamiento del programa y de la instrucción, 2) sobre los alumnos (necesidades y méritos finales) y 3) acerca de la regulación administrativa sobre la calidad del sistema, profesores, organización, etc. De esta forma, Cronbach abre el campo conceptual y funcional de la evaluación educativa mucho más allá del marco conceptual dado por Tyler, aunque en su línea de sugerencias.
- b) La evaluación que se usa para *mejorar un programa mientras este se está aplicando* contribuye más al desarrollo de la educación que la evaluación usada para estimar el valor del producto de un programa ya concluido.
- c) Poner en cuestión la necesidad de que los estudios evaluativos sean de tipo comparativo. Entre las objeciones a esta clase de estudios, el autor destaca el hecho de que, con frecuencia, las diferencias entre las puntuaciones promedio entre-grupos son menores que las intra-grupos, así como otras relativas a las dificultades técnicas que en el marco educativo presentan los diseños comparativos. Cronbach aboga por unos criterios de comparación de tipo absoluto, reclamando la necesidad de una evaluación con referencia al criterio, al defender la valo-

ración con relación a unos objetivos bien definidos, y no la comparación con otros grupos.

- d) Se ponen en cuestión los estudios a gran escala, puesto que las diferencias entre los tratamientos pueden ser muy grandes e impedir que se distingan con claridad las causas de los resultados. Se defienden los estudios más analíticos, bien controlados, que pueden usarse para comparar versiones alternativas de un programa.
- e) Metodológicamente, Cronbach propone que la evaluación debe incluir: 1) estudios de proceso —hechos que tienen lugar en el aula—; 2) medidas de rendimiento y actitudes —cambios observados en los alumnos—; y 3) estudios de seguimiento, esto es, el camino posterior seguido por los estudiantes que han participado en el programa.
- f) Desde esta óptica, las técnicas de evaluación no pueden limitarse a los tests de rendimiento. Los cuestionarios, las entrevistas, la observación sistemática y no sistemática, las pruebas de ensayo, según el autor, ocupan un lugar importante en la evaluación, en contraste al casi exclusivo uso que se hacía de los tests como técnicas de recogida de información.

Si estas reflexiones de Cronbach fueron impactantes, no lo fueron menos las del ensayo de Scriven (1967). Sus fecundas distinciones terminológicas ampliaron enormemente el campo semántico de la evaluación, a la vez que clarificaron el quehacer evaluativo. Destacamos a continuación las aportaciones más significativas:

- a) Se establece de forma tajante la diferencia entre la evaluación como actividad metodológica, lo que el autor llama *meta de la evaluación*, y las funciones de la evaluación en un contexto particular. Así, la eva-

luación como actividad metodológica es esencialmente igual, sea lo que fuera lo que estemos evaluando. El objetivo de la evaluación es invariante, supone, en definitiva, el proceso por el cual estimamos el valor de algo que se evalúa, mientras que las funciones de la evaluación pueden ser enormemente variadas. Estas funciones se relacionan con el uso que se hace de la información recogida.

- b) Scriven señala dos funciones distintas que puede adoptar la evaluación: la formativa y la sumativa. Propone el término de *evaluación formativa* para denominar aquel proceso de evaluación al servicio de un programa en desarrollo con objeto de mejorarlo, y el término de *evaluación sumativa* para aquel proceso orientado a comprobar la eficacia del programa y tomar decisiones sobre su continuidad.
- c) Otra importante contribución de Scriven es la crítica al énfasis que la evaluación da a la consecución de objetivos previamente establecidos, porque, si los objetivos carecen de valor, no tiene ningún interés saber hasta qué punto se han conseguido. Resalta la necesidad de que la evaluación incluya tanto la evaluación de los propios objetivos como el determinar el grado en que estos han sido alcanzados (Scriven, 1973 y 1974).
- d) Clarificadora es también la distinción que hace Scriven entre evaluación intrínseca y evaluación extrínseca, como dos formas diferentes de valorar un elemento de la enseñanza. En una evaluación intrínseca se valora el elemento por sí mismo, mientras que en la evaluación extrínseca se valora el elemento por los efectos que produce en los alumnos. Esta distinción resulta muy importante a la hora de considerar el criterio a utilizar, pues

en la evaluación intrínseca el criterio no se formula en términos de objetivos operativos, mientras que sí se hace en la evaluación extrínseca.

- e) Scriven adopta una posición contraria a Cronbach, defendiendo el carácter comparativo que deben presentar los estudios de evaluación. Admite con Cronbach los problemas técnicos que los estudios comparativos entrañan y la dificultad de explicar las diferencias entre programas, pero considera que la evaluación, como opuesta a la mera descripción, implica emitir un juicio sobre la superioridad o inferioridad de lo que se evalúa con respecto a sus competidores o alternativas.

Estas dos aportaciones comentadas influyeron decisivamente en la comunidad de evaluadores, incidiendo no solo en estudios en la línea de la investigación evaluativa, a la que se referían preferentemente, sino también en la evaluación orientada al sujeto, en la línea de evaluación como *assessment*. Estamos ante la denominada *tercera generación* de la evaluación, que, según Guba y Lincoln (1989), se caracteriza por introducir la valoración, *el juicio*, como un contenido intrínseco en la evaluación. Ahora el evaluador no solo analiza y describe la realidad, además, la valora, la juzga con relación a distintos criterios.

#### **5.4. Otras aportaciones de los sesenta**

Durante estos años sesenta aparecen muchas otras aportaciones que van perfilando una nueva concepción evaluativa que terminará de desarrollarse y, sobre todo, de extenderse en las décadas posteriores. Se percibe que el núcleo conceptual de la evaluación lo constituye la valoración del cambio ocurrido en el alumno como efecto de una situación educativa sistemática, siendo unos

objetivos bien formulados el mejor criterio para valorar este cambio. Así mismo, se comienza a prestar atención no solo a los resultados pretendidos, sino también a los efectos laterales o no pretendidos, e incluso a resultados o efectos a largo plazo (Glaser, 1963; Stake, 1967).

Suchman (1967) profundiza en la convicción de que la evaluación debe basarse en datos objetivos que sean analizados con metodología científica, matizando que la investigación científica es preferentemente teórica y, en cambio, la investigación evaluativa es siempre aplicada. Su principal propósito es descubrir la eficacia, éxito o fracaso de un programa al compararlo con los objetivos propuestos, y, así, trazar las líneas de su posible redefinición. Esta investigación evaluativa, para Suchman, debe tener en cuenta: *a)* la naturaleza del destinatario del objetivo y la del propio objetivo; *b)* el tiempo necesario para que se realice el cambio propuesto; *c)* el conocimiento de si los resultados esperados son dispersos o concentrados; y *d)* los métodos que han de emplearse para alcanzar los objetivos. Suchman, además, defiende la utilización de evaluadores externos para evitar todo tipo de tergiversación de los profesores muy implicados en los procesos instruccionales.

El énfasis en los objetivos y su medida traerá también la necesidad de una nueva orientación a la evaluación, la denominada *evaluación de referencia criterial*. La distinción introducida por Glaser (1963) entre mediciones referidas a normas y criterios tendrá eco al final de la década de los sesenta, precisamente como resultado de las nuevas exigencias que a la evaluación educativa se le planteaban. Así, por ejemplo, cuando Hambleton (1985) estudia las diferencias entre tests referidos al criterio y tests referidos a la norma, señala para los primeros, además de los conocidos objetivos de describir la ejecución del sujeto y

tomar decisiones sobre si domina o no domina un contenido, otro objetivo, como es el de valorar la eficacia de un programa.

Vemos que en el desarrollo y las prácticas evaluativas de esta década de los sesenta se observan dos niveles de actuación. Podemos denominar un nivel el de la *evaluación orientada hacia los individuos*, fundamentalmente alumnos y profesores. El otro nivel es el de la *evaluación orientada a la toma de decisiones sobre el «instrumento» o «tratamiento» o «programa» educativo*. Este último nivel, impulsado también por la evaluación de programas en el ámbito social, será la base para la consolidación en el terreno educativo y en el social, en general, de la evaluación de programas y de la investigación evaluativa.

### **5.5. La eclosión de modelos evaluativos y la profesionalización**

A partir de los setenta, durante aproximadamente dos décadas, se desarrolla la investigación evaluativa con la proliferación de toda clase de modelos que inundan el mercado bibliográfico, *modelos de evaluación* que expresan la propia óptica del autor que los propone sobre *qué es y cómo* debe conducirse un proceso evaluativo. Se trata, por tanto, de una época caracterizada por la pluralidad conceptual y metodológica. Guba y Lincoln (1982) nos hablan de más de cuarenta modelos propuestos en estos años, que enriquecerán considerablemente el vocabulario y la metodología evaluativa. Sin embargo, siempre se ha mantenido el debate sobre si estos supuestos modelos lo eran en el sentido estricto o constituían simplemente aproximaciones diferenciadas. Nosotros nos hemos atrevido a hablar de enfoques modélicos, más que de modelos, puesto que es cada evaluador el que termina constru-

yendo su propio modelo en cada investigación evaluativa, en función del tipo de trabajo y las circunstancias (Escudero, 1993).

Algunos de estos modelos consideran el proceso de evaluación al servicio de las instancias que deben tomar decisiones. Ejemplos notables de ellos son: probablemente el más famoso y utilizado de todos, el CIPP (contexto, input, proceso y producto), propuesto por Stufflebeam y otros (1970), y el CES (toma sus siglas del Centro de la Universidad de California para el Estudio de la Evaluación), de Alkin (1969). La aportación conceptual y metodológica de estos modelos es valorada positivamente entre la comunidad de evaluadores (Popham, 1980; Guba y Lincoln, 1982; House, 1989). Estos autores van más allá de la evaluación centrada en resultados finales, puesto que en sus propuestas suponen diferentes tipos de evaluación, según las necesidades de las decisiones a las que sirven.

Un segundo bloque en la proliferación de modelos es el representado por los *modelos alternativos*, que, con diferentes concepciones de la evaluación y de la metodología a seguir, comienzan a aparecer en la segunda mitad de la década de los setenta. Entre ellos destacan: la *Evaluación Responsable* de Stake (1975), a la que se adhieren Guba y Lincoln (1982), la *Evaluación Democrática* de MacDonald (1976), la *Evaluación Iluminativa* de Parlett y Hamilton (1977) y la *Evaluación como crítica artística* de Eisner (1985).

En líneas generales, este segundo grupo de modelos enfatiza el papel de la *audiencia* de la evaluación y de la relación del evaluador con ella. La audiencia prioritaria de la evaluación en estos modelos no es quien debe tomar las decisiones, como en los modelos orientados a la toma de decisiones, ni el responsable de elaborar los

currículos u objetivos, como en los modelos de consecución de metas. La audiencia prioritaria son los propios participantes del programa. La relación entre el evaluador y la audiencia, en palabras de Guba y Lincoln (1982), debe ser «transaccional y fenomenológica». Se trata de modelos que propugnan una evaluación de tipo etnográfico, de ahí que la metodología que consideran más adecuada sea la propia de la antropología social (Parlett y Hamilton, 1977).

Este proceso de eclosión de modelos condujo a un amplio abanico de concepciones teóricas y metodológicas alrededor del concepto de evaluación. Ello explica que, cuando Nevo (1989) pretende realizar una conceptualización de la evaluación, a partir de la revisión de la literatura especializada, atendiendo a los tópicos qué es la evaluación, qué funciones tiene, cuál es el objeto de evaluación..., no encuentre una única respuesta a estas cuestiones. Es fácilmente comprensible que las exigencias que plantea la evaluación de programas, de una parte, y la evaluación para la toma de decisiones sobre los individuos, de otra, conduzcan a una gran variedad de esquemas evaluativos reales utilizados por profesores, directores, inspectores y administradores públicos. Pero también es cierto que bajo esta diversidad subyacen diferentes concepciones teóricas y metodológicas sobre la evaluación. Diferentes concepciones que han dado lugar a una apertura y pluralidad conceptual en el ámbito de la evaluación en varios sentidos. A continuación destacamos los puntos más sobresalientes de esta pluralidad conceptual (Escudero, 2003):

- a) *Diferentes conceptos de evaluación.* Por una parte, existe la clásica definición dada por Tyler: *la evaluación como el proceso de determinar el grado de congruencia entre las realizaciones y los objetivos previamente estable-*

*cidos*, a la que corresponden los modelos orientados hacia la consecución de metas. Contrasta esta definición con aquella más amplia que se propugna desde los modelos orientados a la toma de decisiones: *la evaluación como el proceso de determinar, obtener y proporcionar información relevante para juzgar decisiones alternativas*, defendida por Alkin (1969), Stufflebeam y otros (1970), MacDonald (1976) y Cronbach (1982).

Además, el concepto de evaluación de Scriven (1967), como el proceso de estimar el valor o el mérito de algo, es retomado por Cronbach (1982), Guba y Lincoln (1982) y House (1989), con objeto de señalar las diferencias que comportarían los juicios valorativos en caso de estimar el *mérito* (se vincularía a características intrínsecas de lo que se evalúa) o el *valor* (se vincularía al uso y aplicación que tendría para un contexto determinado).

- b) *Diferentes criterios*. De las definiciones apuntadas anteriormente se desprende que el criterio a utilizar para la valoración de la información también cambia. Desde la óptica de la consecución de metas, una buena y operativa definición de los objetivos constituye el criterio fundamental. Desde la perspectiva de las decisiones y situados dentro de un contexto político, Stufflebeam y colaboradores, Alkin y MacDonald llegan a sugerir incluso la no valoración de la información por parte del evaluador, siendo el que toma las decisiones el responsable de su valoración.

Las definiciones de evaluación que acentúan la determinación del «mérito» como objetivo de la evaluación utilizan criterios estándar sobre los que los expertos o profesionales están de acuerdo. Se trata

de modelos relacionados con la acreditación y el enjuiciamiento profesional (Popham, 1980).

Los autores (Stake, 1975; Parlett y Hamilton, 1977; Guba y Lincoln, 1982) que acentúan el proceso de evaluación al servicio de determinar el «valor», más que el «mérito» de la entidad u objeto evaluado, abogan por que el criterio de valoración fundamental sean las necesidades contextuales en las que esta se inserta. Así, Guba y Lincoln enuncian los términos de la comparación valorativa: de un lado, las características del objeto evaluado; y, de otro, las necesidades, expectativas y valores del grupo al que afecta.

- c) *Pluralidad de procesos evaluativos*, dependiendo de la percepción teórica que sobre la evaluación se mantenga. Los modelos de evaluación citados y otros más que pueden encontrarse en la bibliografía representan diferentes propuestas para conducir una evaluación.
- d) *Pluralidad de objetos de evaluación*. Como dice Nevo (1989), existen dos conclusiones importantes que se obtienen de la revisión de la bibliografía sobre la evaluación. Por un lado, cualquier cosa puede ser objeto de evaluación y esta no debería limitarse a estudiantes y profesores; por otro, una clara identificación del objeto de evaluación es una parte importante en cualquier diseño de evaluación.
- e) *Apertura*, reconocida en general por todos los autores, de la información necesaria en un proceso evaluativo para dar cabida no solo a los resultados pretendidos, sino a los *efectos posibles* de un programa, sean pretendidos o no. Incluso Scriven (1973 y 1974) propone una evaluación en la que no se ten-

gan en cuenta los objetivos pretendidos, sino que se valoren todos los efectos posibles. Apertura también respecto a la recogida de información no solo del producto final, sino también del *proceso*. Y apertura en la consideración de diferentes resultados de *corto y largo alcance*. Por último, apertura también en considerar no solo resultados de tipo cognitivo, sino también afectivos (Anderson y Ball, 1983).

- f) Pluralidad también reconocida de las *funciones* de la evaluación, recogiendo la propuesta de Scriven de evaluación formativa y sumativa, y añadiéndose otras de tipo sociopolítico y administrativas (Nevo, 1989).
- g) Diferencias en el papel desempeñado por *el evaluador*; lo que ha venido a llamarse *evaluación interna* vs. *evaluación externa*. No obstante, una relación directa entre el evaluador y las diferentes audiencias de la evaluación es reconocida por la mayoría de los autores (Nevo, 1989; Weiss, 1983; Rutman, 1984).
- h) *Pluralidad de audiencias* de la evaluación y, por consiguiente, *pluralidad en los informes* de evaluación. Desde informes narrativos, informales, hasta informes muy estructurados (Anderson y Ball, 1983).
- i) *Pluralidad metodológica*. Las cuestiones metodológicas surgen desde la dimensión de la evaluación como investigación evaluativa, que viene definida en gran medida por la diversidad metodológica.

El anterior resumen recoge las aportaciones a la evaluación en los años setenta y ochenta, la denominada *época de la profesionalización* (Stufflebeam y Shinkfield, 1987; Madaus y otros, 1991; Mateo y otros, 1993), en la que, además de los innumerables modelos de los setenta, se profundizó en los planteamientos teóricos y prácticos y se consolidó la evaluación como *investigación evaluativa*

en los términos antes definidos. En este contexto, lógicamente, aparecen muchas nuevas revistas especializadas como *Educational Evaluation and Policy Analysis*, *Studies in Educational Evaluation*, *Evaluation Review*, *New Directions for Program Evaluation*, *Evaluation and Program Planning*, *American Journal of Evaluation*, *Evaluation News*, *Evaluation Studies Review Annual*, *Journal of Personnel Evaluation in Education...*, se fundan asociaciones científicas relacionadas con el desarrollo de la evaluación, y las universidades empiezan a ofrecer cursos y programas de investigación evaluativa, no solo en postgrados y programas de doctorado, sino también en planes de estudio para titulaciones de grado. En paralelo, muchas de las principales universidades americanas crean centros de investigación en evaluación (The Oral History Project Team, 2008), así, por ejemplo: el Stanford University Evaluation Consortium, liderado por Cronbach; el Center for the Study of Evaluation en UCLA, dirigido por Alkin; el CIRCE en Illinois, liderado por Hastings y Stake; el Center for Evaluation de Minnesota, dirigido por Welch y Merwin; el Boston College Center for Testing and Evaluation, dirigido por Madaus; el Evaluation Research Center de la Universidad de Virginia, con Provus y Covert; el Evaluation Institute de la Universidad de San Francisco, dirigido por Scriven; y el Evaluation Center de Stufflebeam, inicialmente en la Ohio State University y posteriormente en la Western Michigan University.

## **5.6. La cuarta generación**

A finales de los ochenta, tras todo este desarrollo antes descrito, Guba y Lincoln (1989) ofrecen una alternativa evaluadora que denominan *cuarta generación*, pretendiendo superar lo que, según estos autores, son defi-

ciencias de las tres generaciones anteriores, tales como una visión gestora de la evaluación, una escasa atención al pluralismo de valores y un excesivo apego al paradigma positivista. A su alternativa Guba y Lincoln la denominan *respondente y constructivista*, integrando de alguna manera el enfoque *respondente* propuesto en primer lugar por Stake (1975) y la epistemología *postmoderna* del constructivismo (Russell y Willinsky, 1997). Las demandas, las preocupaciones y los asuntos de los implicados y responsables (*stakeholders*) sirven como foco organizativo de la evaluación (como base para determinar qué información se necesita), que se lleva a cabo dentro de los planteamientos metodológicos del paradigma constructivista.

La utilización de las demandas, preocupaciones y asuntos de los implicados es necesaria, según Guba y Lincoln, porque:

- a) Son *grupos de riesgo* ante la evaluación, y sus problemas deben ser convenientemente contemplados, de manera que se sientan protegidos ante tal riesgo.
- b) Los resultados pueden ser utilizados en su *contra* en diferentes sentidos, sobre todo si están al margen del proceso.
- c) Son potenciales usuarios de la información resultante de la evaluación.
- d) Pueden ampliar y mejorar el rango de la evaluación.
- e) Se produce una interacción positiva entre los distintos implicados.

El cambio paradigmático lo justifican estos autores porque:

- a) La metodología convencional no contempla la necesidad de identificar las demandas, preocupaciones y asuntos de los implicados.

- b) Para llevar a cabo lo anterior se necesita una postura de descubrimiento, más que de verificación, típica del positivismo.
- c) No se tienen en cuenta suficientemente los factores contextuales.
- d) No se proporcionan medios para valoraciones caso por caso.
- e) La supuesta neutralidad de la metodología convencional es de dudosa utilidad cuando se buscan juicios de valor acerca de un objeto social.

Partiendo de estas premisas, el evaluador es responsable de determinadas tareas, que realizará secuencialmente o en paralelo, construyendo un proceso ordenado y sistemático de trabajo. Las responsabilidades básicas del evaluador de la cuarta generación son las siguientes:

- 1) Identificar todos los implicados con riesgo en la evaluación.
- 2) Resaltar para cada grupo de implicados sus construcciones acerca de lo evaluado y sus demandas y preocupaciones al respecto.
- 3) Proporcionar un contexto y una metodología hermenéutica para poder tener en cuenta, comprender y criticar las diferentes construcciones, demandas y preocupaciones.
- 4) Generar el máximo acuerdo posible acerca de dichas construcciones, demandas y preocupaciones.
- 5) Preparar una agenda para la negociación acerca de temas no consensuados.
- 6) Recoger y proporcionar la información necesaria para la negociación.
- 7) Formar y hacer de mediador para un fórum de implicados para la negociación.
- 8) Desarrollar y elaborar informes para cada grupo de implicados sobre los distintos acuerdos y reso-

luciones acerca de los intereses propios y de los de otros grupos.

- 9) Reciclar la evaluación siempre que queden asuntos pendientes de resolución.

Para finalizar este sucinto análisis de la cuarta generación de Guba y Lincoln, recogemos los rasgos que, según ellos, mejor definen a la investigación evaluativa. Se trata de un proceso: *sociopolítico*, de *colaboración*, de *enseñanza/aprendizaje continuo*, *recursivo* y *altamente divergente*, *emergente*, con resultados *impredecibles* y que *crea realidad*.

En esta evaluación se retienen las características del evaluador fruto de las tres primeras generaciones, esto es, la de *técnico*, la de *analista* y la de *juez*, pero estas deben ampliarse con destrezas para recoger e interpretar datos cualitativos (Patton, 1984), como la de historiador e iluminador, la de mediador de juicios, así como un papel más activo como evaluador en un contexto sociopolítico concreto.

### **5.7. Las recomendaciones de Stufflebeam**

Este recorrido analítico-histórico de grandes impulsos en la formalización de la actual investigación evaluativa debe recoger necesariamente las ideas y recomendaciones que ya más recientemente nos ha venido ofreciendo una de las figuras señeras de este campo en la segunda mitad del siglo XX y principios del XXI. Nos estamos refiriendo a Daniel L. Stufflebeam, que propuso el modelo CIPP (el más utilizado) a finales de los sesenta, que de 1975 a 1988 fue presidente del Joint Committee on Standards for Educational Evaluation y posteriormente director del Evaluation Center de la Western Michigan University (sede del Joint Committee) y del CREATE (Center

for Research on Educational Accountability and Teacher Evaluation), centro auspiciado y financiado por el Departamento de Educación del Gobierno estadounidense.

Recogiendo estos planteamientos (Stufflebeam, 1994, 1998, 1999, 2000 y 2001), en los que se han ido integrando ideas de diversos evaluadores también notables, ofrecemos una de las últimas grandes aportaciones a la actual concepción de la investigación evaluativa.

Se parte de los cuatro principios del Joint Committee (1981 y 1988), esto es, de la idea de que cualquier buen trabajo de investigación evaluativa debe ser: a) *útil*, esto es, proporcionar información a tiempo e influir; b) *factible*, esto es, debe suponer un esfuerzo razonable y debe ser políticamente viable; c) *apropiado, adecuado, legítimo*, esto es, ético y justo con los implicados; y d) *seguro y preciso* a la hora de ofrecer información y juicios sobre el objeto de la evaluación. Además, la evaluación se ve como una «transdisciplina», pues es aplicable a muchas disciplinas diferentes y a muchos objetos diversos (Scriven, 1994).

Stufflebeam apela a la responsabilidad del evaluador, que debe actuar de acuerdo con principios aceptados por la sociedad y con criterios de profesionalidad, emitir juicios sobre la calidad y el valor del objeto evaluado, y debe asistir a los implicados en la interpretación y utilización de su información y sus juicios. Sin embargo, es también su deber, y su derecho, estar al margen de la lucha y la responsabilidad política por la toma de decisiones y por las decisiones tomadas.

Para evaluar la educación en una sociedad moderna, Stufflebeam (1994) nos dice que se deben tomar algunos criterios básicos de referencia, como los siguientes:

- Las *necesidades educativas*. Es necesario preguntarse si la educación que se proporciona cubre las necesidades de los estudiantes y de sus familias en todos

los terrenos a la vista de los derechos básicos; en este caso, dentro de una sociedad democrática.

- La *equidad*. Hay que preguntarse si el sistema es justo y equitativo a la hora de proporcionar servicios educativos, el acceso a los mismos, la consecución de metas, el desarrollo de aspiraciones y la cobertura para todos los sectores de la comunidad.
- La *factibilidad*. Hay que cuestionar la eficiencia en la utilización y distribución de recursos, la adecuación y viabilidad de las normas legales, el compromiso y participación de los implicados y todo lo que hace que el esfuerzo educativo produzca el máximo de frutos posibles.
- La *excelencia* como objetivo permanente de búsqueda. La mejora de la calidad, a partir del análisis de las prácticas pasadas y presentes, es uno de los fundamentos de la investigación evaluativa.

Tomando el referente de estos criterios y sus derivaciones, Stufflebeam enumera una serie de recomendaciones para llevar a cabo buenas investigaciones evaluativas y mejorar el sistema educativo. Estas recomendaciones son las siguientes:

- 1) Los planes de evaluación deben satisfacer los cuatro requerimientos de *utilidad, factibilidad, legitimidad y precisión* (Joint Committee, 1981 y 1988).
- 2) Las entidades educativas deben examinarse por su integración y servicio a los *principios de la sociedad democrática*, equidad, bienestar, etc.
- 3) Las entidades educativas deben ser valoradas tanto por su *mérito* (valor intrínseco, calidad respecto a criterios generales) como por su *valor* (valor extrínseco, calidad y servicio para un contexto particular), como por su *significación* en la

realidad del contexto en el que se ubican. Scriven (1998) nos señala que, usando otras denominaciones habituales, *mérito* tiene bastante equivalencia con el término *calidad*, *valor* con el de *relación coste-eficacia* y *significación* con el de *importancia*. En todo caso, los tres conceptos son dependientes del contexto, sobre todo *significación*, de manera que entender la diferencia entre dependencia del contexto y arbitrariedad es parte de la comprensión de la lógica de la evaluación.

- 4) La evaluación de profesores, instituciones educativas, programas, etc., debe relacionarse siempre con el conjunto de sus deberes, responsabilidades y obligaciones profesionales o institucionales, etc. Quizás uno de los retos que deben abordar los sistemas educativos es la definición más clara y precisa de estos *deberes y responsabilidades*. Sin ello, la evaluación es problemática, incluso en el terreno formativo (Scriven, 1991a).
- 5) Los estudios evaluativos deben ser capaces de valorar en qué medida los profesores y las instituciones educativas son *responsables y rinden cuentas* del cumplimiento de sus deberes y obligaciones profesionales (Scriven, 1994).
- 6) Los estudios evaluativos deben proporcionar *direcciones para la mejora*, porque no basta con emitir un juicio sobre el mérito o el valor de algo.
- 7) Recogiendo los puntos anteriores, todo estudio evaluativo debe tener un componente *formativo* y otro *sumativo*.
- 8) Se debe promover la *autoevaluación* profesional, proporcionando a los educadores las destrezas para ello y favoreciendo actitudes positivas hacia ella (Madaus y otros, 1991).

- 9) *La evaluación del contexto* (necesidades, oportunidades, problemas en un área...) debe emplearse de manera *prospectiva* para localizar bien las metas y objetivos y definir prioridades. Así mismo, la evaluación del contexto debe utilizarse *retrospectivamente*, para juzgar bien el valor de los servicios y resultados educativos, en relación con las necesidades de los estudiantes (Madaus y otros, 1991; Scriven, 1991).
- 10) *La evaluación de las entradas (inputs)* debe emplearse de manera *prospectiva* para asegurar el uso de un rango adecuado de enfoques según las necesidades y los planes.
- 11) *La evaluación del proceso* debe usarse de manera *prospectiva* para mejorar el plan de trabajo, pero también de manera *retrospectiva* para juzgar hasta qué punto la calidad del proceso determina el porqué los resultados son de un nivel u otro (Stufflebeam y Shinkfield, 1987).
- 12) *La evaluación del producto* es el medio para identificar los resultados buscados y no buscados en los participantes o afectados por el objeto evaluado. Se necesita una valoración *prospectiva* de los resultados para orientar el proceso y detectar zonas de necesidades. Se necesita una evaluación *retrospectiva* del producto para poder juzgar en conjunto el mérito y el valor del objeto evaluado.
- 13) Los estudios evaluativos se deben apoyar en la *comunicación* y en la *inclusión* sustantiva y funcional de los implicados (*stakeholders*) con las cuestiones claves, criterios, hallazgos e implicaciones de la evaluación, así como en la promoción de la aceptación y el uso de sus resultados (Chelimsky, 1998). Más aún, los estudios evaluativos deben

conceptualizarse y utilizarse sistemáticamente como parte del proceso de mejora educativa a largo plazo (Stronge y Helm, 1991; Keefe, 1994) y de fundamento para la acción contra las discriminaciones sociales (Mertens, 1999). La *evaluación para el desarrollo o evaluación capacitadora (empowerment evaluation)*, que defiende Fetterman (2000), es un procedimiento, de base democrática, de participación de los implicados en el programa evaluado, para promover la autonomía de los mismos en la resolución de sus problemas. Weiss (1998) nos alerta de que la evaluación participativa incrementa la probabilidad de que se utilicen los resultados de la evaluación, pero también la de que sea conservadora en su concepción, pues es difícil pensar que los responsables de una organización pongan en cuestión el fundamento y el sistema de poder de la misma. Generalmente, su interés es el cambio de cosas pequeñas.

- 14) Los estudios evaluativos deben emplear *múltiples perspectivas, múltiples medidas de resultados*, y métodos tanto *cuantitativos* como *cualitativos* para recoger y analizar la información. La pluralidad y complejidad del fenómeno educativo hace necesario emplear enfoques múltiples y multidimensionales en los estudios evaluativos (Scriven, 1991).
- 15) Los estudios evaluativos deben ser evaluados, incluyendo *metaevaluaciones formativas* para mejorar su calidad y su uso y *metaevaluaciones sumativas* para ayudar a los usuarios en la interpretación de sus hallazgos y proporcionar sugerencias para mejorar futuras evaluaciones.

Estas quince recomendaciones proporcionan elementos esenciales para un enfoque de los estudios evaluativos

que Stufflebeam denomina *objetivista* y que se basa en la teoría ética de que la bondad moral es objetiva e independiente de los sentimientos personales o meramente humanos.

Sin entrar en el debate sobre estas valoraciones finales de Stufflebeam, ni en análisis comparativos con otras propuestas, por ejemplo con las de Guba y Lincoln (1989), nos resulta evidente que las concepciones de la investigación evaluativa son diversas, dependiendo del origen epistemológico desde el que se parte, pero apareciendo claros y contundentes algunos elementos comunes a todas las perspectivas, como la *contextualización*, el *servicio a la sociedad*, la *diversidad metodológica*, la *atención, respeto y participación de los implicados*, etc., así como una mayor *profesionalización* de los evaluadores y una mayor *institucionalización* de los estudios (Worthen y Sanders, 1991).

El propio Stufflebeam (1998) reconoce el conflicto de los planteamientos del Joint Committee on Standards for Educational Evaluation con las posiciones de la corriente evaluadora denominada *postmodernista*, representada, entre otros, por Guba y Lincoln y Stake, pero no acepta que existan razones para actitudes de escepticismo y frustración con las prácticas evaluadoras actuales, porque se dan muchos espacios de aproximación y el desarrollo de estándares de evaluación es perfectamente compatible con la atención a los diversos implicados, valores, contextos sociales y métodos. Stufflebeam defiende una mayor colaboración en la mejora de las evaluaciones, estableciendo los estándares de manera participativa, pues cree que es posible la aproximación de planteamientos, con contribuciones importantes desde todos los puntos de vista.

Weiss (1998) también toma posiciones parecidas cuando nos dice que las ideas constructivistas deben hacernos pensar más cuidadosamente al usar los resulta-

dos de las evaluaciones, sintetizarlas y establecer generalizaciones, pero duda que todo haya que interpretarlo en términos exclusivamente individuales, pues existen muchos elementos comunes entre las personas, los programas y las instituciones.

## **5.8. El auge de la evaluación institucional**

No queremos terminar este sucinto recorrido de hitos en la construcción de la investigación evaluativa desde el terreno de la educación sin referirnos al auge que ha adquirido en las últimas décadas la *evaluación institucional* en el ámbito de la educación superior, en la gran mayoría de los sistemas universitarios desarrollados y, en particular, en el contexto europeo (Escudero, 2007).

La evaluación institucional sintetiza y recoge la gran mayoría de objetos de interés y de ámbitos de actuación de la investigación evaluativa, políticas, programas, sistemas organizativos y de gobierno, relaciones internas y externas, recursos personales y materiales, infraestructuras, etc., englobando e integrando además los principios y métodos defendidos desde el campo de la calidad institucional y de los procesos de su control, aseguramiento y mejora.

Esta realidad ha impulsado y fortalecido la utilización de ciertas metodologías en este campo de investigación, tales como, por ejemplo, el desarrollo de autoinformes, la utilización de indicadores de rendimiento y los contrastes entre visiones internas y externas (Escudero, 2002).

Al hilo de este impulso se han potenciado los enfoques y organismos de evaluación externa (De Miguel, Escudero y Rodríguez, 1998), creándose agencias y oficinas intermedias, y redes de carácter regional, nacional e internacional, de apoyo a las evaluaciones y procesos de acreditación y certificación de calidad.



SUCINTA SÍNTESIS SOBRE LA CUESTIÓN  
DE LOS MODELOS Y ENFOQUES  
DE LA INVESTIGACIÓN EVALUATIVA

Tras el anterior análisis, parece oportuno, a modo de síntesis, recoger y resaltar los que son considerados principales modelos, planteamientos metodológicos, diseños, perspectivas y visiones de la evaluación en la actualidad. Se trata de un complemento necesario para una visión como la histórica, que, por su linealidad, corre el riesgo de ofrecer una imagen disciplinar artificiosamente fraccionada.

Hemos visto que en la década de los setenta y en sus alrededores se produce una especie de eclosión de propuestas evaluativas que tradicionalmente han venido siendo denominadas *modelos* (Castillo y Gento, 1995) y en algunos casos *diseños* (Arnal y otros, 1992) de investigación evaluativa. Para nosotros, quizás sea el término *enfoque* evaluativo el más apropiado, aunque aceptemos seguir hablando de modelos y diseños por simple tradición académica.

Sabemos que existieron varias decenas de estas propuestas, pero muy concentradas en el tiempo, en la década citada. De hecho, el asunto de los modelos propuestos para la evaluación parece un tema prácticamente cerrado desde hace varios lustros. Ya no surgen nuevos modelos o propuestas, salvo alguna excepción, como veremos

más adelante. A pesar de lo dicho, se sigue hablando de modelos, métodos y diseños en la literatura especializada, sobre todo buscando su clasificación de acuerdo con diversos criterios, origen paradigmático, propósito, metodología, etc. También en las clasificaciones, no solo en los modelos, existe diversidad, lo que prueba que, además de dinamismo académico en el terreno de la investigación evaluativa, todavía existen cuestiones teóricas abiertas.

Nuestra idea es que, a la hora de plantearnos una investigación evaluativa, no contamos con unos pocos modelos bien fundamentados, definidos, estructurados y completos, entre los que elegir uno de ellos, sino que tenemos distintos enfoques modélicos y un amplio soporte teórico, metodológico y técnico, que permiten al evaluador ir respondiendo de manera adecuada a las distintas cuestiones que le va planteando el proceso de investigación, ayudándole a configurar un *plan global*, un *organigrama coherente*, un «modelo» científicamente robusto para llevar a cabo su evaluación (Escudero, 1993). ¿Cuáles son las cuestiones que hay que responder en este proceso de construcción modélica? Apoyándonos en las aportaciones de diferentes autores (Worthen y Sanders, 1973; Nevo, 1989; Smith y Haver, 1990), deben responderse y delimitar su respuesta al construir un modelo de investigación evaluativa los aspectos siguientes:

- 1) Objeto de la investigación evaluativa.
- 2) Propósito, objetivos.
- 3) Audiencias/implicados/clientela.
- 4) Énfasis/aspectos prioritarios o preferentes.
- 5) Criterios de mérito o valor.
- 6) Información a recoger.
- 7) Métodos de recogida de información.
- 8) Métodos de análisis.

- 9) Agentes del proceso.
- 10) Secuenciación del proceso.
- 11) Informes/utilización de resultados.
- 12) Límites de la evaluación.
- 13) Evaluación de la propia investigación/metaevaluación.

Para definir estos elementos hay que buscar, lógicamente, el apoyo de los diferentes enfoques modélicos, métodos, procedimientos, etc., que la investigación evaluativa ha desarrollado.

Scriven (1994) identifica *seis visiones* o enfoques alternativos en la fase «explosiva» de los modelos, además de algunas más que denomina «exóticas» y que se mueven entre los modelos de jurisprudencia y de experto. A continuación comentamos sucintamente estas visiones y los «modelos» que se adscriben a ellas.

*La visión fuerte hacia la toma de decisiones* (Visión A) concibe al evaluador investigando con el objetivo de llegar a conclusiones evaluativas que ayuden al que debe tomar decisiones. Los que apoyan este enfoque se preocupan de si el programa alcanza sus objetivos, pero van más allá, cuestionándose si tales objetivos cubren las necesidades que deben cubrir. Esta posición es mantenida, aunque no la hiciera explícita, por Ralph Tyler y aparece extensamente elaborada en el modelo CIPP (Stufflebeam y otros, 1970).

Según el planteamiento tyleriano, las decisiones acerca de un programa deben basarse en el grado de coincidencia entre los objetivos y los resultados. A diferencia de Tyler, Stufflebeam ofrece una perspectiva más amplia de los contenidos a evaluar. Estos son las cuatro dimensiones que identifican su modelo: *contexto* (C), donde tiene lugar el programa o está la institución; *inputs* (I), elementos y recursos de partida; *proceso* (P), que hay que

seguir hacia la meta; y *producto* (P) que se obtiene. Además, se deja constancia de que el objetivo primordial de la investigación evaluativa es la mejora, la toma de decisiones para la mejora de todas y cada una de las cuatro dimensiones antes citadas.

*La visión débil hacia la toma de decisiones* (Visión B) concibe al evaluador proporcionando información relevante para la toma de decisiones, pero no le obliga a emitir conclusiones evaluativas o críticas a los objetivos de los programas. El representante teórico más genuino es Marv Alkin (1969), que define la evaluación como un proceso factual de recogida y generación de información al servicio del que toma las decisiones, pero es este el que tiene que sacar las conclusiones evaluativas. Esta posición es lógicamente popular entre los que piensan que la verdadera ciencia no debe o no puede entrar en cuestiones de juicios de valor. El modelo de Alkin se conoce como CES (Centro para el Estudio de la Evaluación), planteando las siguientes fases: valoración de las necesidades y fijación del problema, planificación del programa, evaluación de la instrumentalización, evaluación de progresos y evaluación de resultados.

*La visión relativista* (Visión C) también mantiene la distancia de las conclusiones evaluativas, pero usando el marco de valores de los clientes, sin un juicio por parte del evaluador acerca de esos valores o alguna referencia a otros. Esta visión y la anterior han sido el camino que ha permitido integrarse sin problemas en el «carro» de la investigación evaluativa a muchos científicos sociales. De hecho, uno de los textos más utilizados de evaluación en el ámbito de las ciencias sociales (Rossi y Freeman, 1993) toma preferentemente esta perspectiva.

Las visiones B y C son las posiciones de los científicos entroncados con una concepción libre de valores de la

ciencia. En cambio, los que participan de la visión A proceden de un paradigma diferente, probablemente debido a su conexión académica con la historia, la filosofía, la educación comparada y la administración educativa.

Alkin (1991) revisó sus planteamientos de dos décadas atrás, pero siguió sin incluir los términos de mérito, valor o valía; termina definiendo un Sistema de Información para la Gestión (Management Information System-MIS) para uso del que toma decisiones, pero no ofrece valoraciones al respecto.

Pero la forma más simple de la *visión relativista* (Visión C) es la desarrollada en el «modelo de discrepancia» de evaluación de Malcolm Provus (1971). Las discrepancias son las divergencias con la secuencia de tareas proyectadas y la temporalización prevista. Este modelo es muy cercano al control de programas en sentido convencional; es una especie de simulación de una evaluación.

*La visión de la descripción fértil, rica, completa* (Visión D) es la que entiende la evaluación como una tarea etnográfica o periodística, en la que el investigador informa de lo que ve sin intentar emitir afirmaciones valorativas o inferir conclusiones evaluativas, ni siquiera en el marco de los valores del cliente, como en la visión relativista. Esta visión ha sido defendida por Robert Stake y muchos de los teóricos británicos. Se trata de una especie de versión naturalista de la visión B, tiene algo de sabor relativista y a veces parece precursora de la visión de la cuarta generación. Se centra en la observación, en lo observable, más que en la inferencia. Recientemente se le ha denominado *visión de la descripción sólida, fuerte*, para evitar el término *rica*, que parece más evaluativo.

Stake, en su primera etapa, es tyleriano en cuanto a concepción evaluativa centrada en los objetivos planteados, proponiendo el método de evaluación de la *figura*

(Stake, 1967) como rostro o imagen total de la evaluación. Esta gira en torno a los tres componentes (*antecedentes, transacciones y resultados*), elaborando dos matrices de datos, una de *descripción* y otra de *juicio*. En la primera se recogen, de un lado, las *intenciones* y, de otro, las *observaciones*; y, en la segunda, las *normas*, lo que se aprueba y los *juicios*, lo que se cree que debe ser. A mediados de los setenta, Stake se aleja de la tradición tyleriana de preocupación por los objetivos y revisa su método de evaluación hacia un planteamiento que él califica como «*respondente*» (Stake, 1975), asumiendo que los objetivos del programa pueden modificarse sobre la marcha, con la finalidad de ofrecer una visión completa y holística del programa y *responder* a los problemas y cuestiones reales que plantean los implicados. Según Stufflebeam y Shinkfield (1987), este modelo hizo de Stake el líder de una nueva escuela de evaluación que exige un método pluralista, flexible, interactivo, holístico, subjetivo y orientado al servicio. Este modelo sugiere la «atención al cliente» propuesta por Scriven (1973), valorando sus necesidades y expectativas.

En el método *respondente* el evaluador ha de entrevistar a los implicados para conocer sus puntos de vista y buscar la confluencia de las diversas perspectivas. El evaluador deberá interpretar las opiniones y diferencias de puntos de vista (Stacher y Davis, 1990) y presentar una amplia gama de opiniones o juicios, en lugar de presentar sus conclusiones personales.

*La visión del proceso social* (Visión E), que cristalizó hace tres décadas alrededor de un grupo de la Universidad de Stanford dirigido por Lee J. Cronbach (1980), resta importancia a la orientación sumativa de la evaluación (decisiones externas sobre los programas y rendición de cuentas), enfatizando la *comprensión*, la *planificación* y la

*mejora* de los programas sociales a los que sirve. Sus posiciones quedaban claramente establecidas en noventa y cinco tesis que han tenido una enorme difusión e influencia entre los investigadores y los usuarios de las evaluaciones.

Ernie House (1989), un teórico y un práctico de la evaluación bastante independiente de corrientes en boga, también marcó el entronque social de los programas, pero se distinguía sobre todo por su énfasis en las dimensiones más éticas y argumentales de la evaluación, quizás motivado por la ausencia de estas facetas en los planteamientos de Cronbach y sus colaboradores.

La *visión constructivista de la cuarta generación* (Visión F) es la última de estas seis visiones que describe Scriven (1994), siendo mantenida por Guba y Lincoln (1989) y seguida por muchos evaluadores americanos y británicos. Esta visión rechaza una evaluación orientada a la búsqueda de calidad, mérito, valor, etc., y favorece la idea de que ello es el resultado de la construcción por individuos y la negociación de grupos. Esto significa, según Scriven, que el conocimiento científico de todo tipo es sospechoso, discutible y no objetivo. Lo mismo ocurre con todo trabajo analítico, como el análisis filosófico, incluido el suyo. Scriven apunta que el propio Guba ha sido siempre consciente de las potenciales «contradicciones» de su posición.

De esta revisión de Scriven quedan al margen algunas posiciones evaluativas tradicionalmente recogidas y tratadas por los analistas. Así, por ejemplo, Suchman (1967) ofrece un diseño evaluativo basado en el *método científico* o, al menos, en alguna variación o adaptación del mismo. Owens (1973) y Wolf (1974 y 1975) proponen un método de *contraposición* o discusión que sobre un programa llevan a cabo dos grupos de evaluadores, partidarios y adversarios, para proporcionar informa-

ción pertinente a quienes toman decisiones. Eisner (1975) plantea la evaluación en términos similares al proceso de crítica artística.

El propio Scriven (1967 y 1973) propuso centrar la evaluación en la *atención al cliente* y no tanto en las metas previstas, puesto que frecuentemente los logros no previstos son más importantes que los que figuran en la planificación del programa. Por ello, se suele denominar a su enfoque *evaluación sin metas*. El evaluador determina el valor o mérito del programa para informar a los usuarios; se trata de algo así como un intermediario informativo.

*La evaluación iluminativa* (Parlett y Hamilton, 1977) tiene un enfoque holístico, descriptivo e interpretativo, con la pretensión de iluminar un complejo rango de cuestiones que se dan de manera interactiva. *La evaluación democrática* de MacDonald (1976) supone la participación colaborativa de los implicados, siendo el contraste de opiniones de los implicados el elemento evaluativo primordial.

Scriven (1994) analiza las seis visiones críticamente y se muestra más cercano a la visión A, *la visión fuerte sobre la toma de decisiones*, representada fundamentalmente por el modelo CIPP de Stufflebeam y sus planteamientos, pues dice que es la más cercana de todas a la *visión del sentido común*, que es la que tienen los evaluadores trabajando con sus programas, de la misma manera que los médicos trabajan con los pacientes, haciéndolo lo mejor posible, independientemente del tipo y del estado general del paciente. Scriven quiere extender esta visión con una visión o perspectiva que denomina *transdisciplinar* y que él califica como significativamente distinta de la aludida visión A y radicalmente diferente de las restantes.

En la *perspectiva transdisciplinar*, la investigación evaluativa tiene dos componentes: el conjunto de campos de

aplicación de la evaluación y el contenido de la propia disciplina. Algo parecido a lo que ocurre en disciplinas como la estadística y la medición. En definitiva, es una disciplina que incluye sus propios contenidos y los de otras muchas disciplinas; su preocupación por el análisis y mejora se extiende a numerosas disciplinas, es transdisciplinar. La investigación evaluativa, manteniendo su idiosincrasia como disciplina autónoma, proporciona herramientas y apoyos básicos a otras disciplinas muy diversas (Scriven, 2003).

Esta visión es *objetivista*, como la A, y defiende que el evaluador determine el mérito o el valor del programa, del personal o de los procesos y productos investigados. En tal sentido, se debe establecer de manera explícita y defender la lógica utilizada en la inferencia de conclusiones evaluativas a partir de las premisas definicionales y factuales. Así mismo, se deben criticar las falacias argumentales de la doctrina libre de valores (Evaluation Thesaurus, 1991).

En segundo lugar, la *perspectiva transdisciplinar* se orienta hacia el consumidor, más que hacia el gestor o intermediario. No se trata de una orientación exclusiva hacia el consumidor, pero se considera que es el consumidor el que justifica el programa y que el bien común tiene la primacía de la evaluación. A partir de aquí, también se produce información valiosa para el gestor, que decide y se pueden analizar los productos de un programa o institución a la vista de sus objetivos. Esta posición no solo ve legitimidad en la emisión de conclusiones evaluativas por parte del investigador, sino que cree necesario hacerlo en la gran mayoría de las ocasiones.

Se trata también de una *visión generalizada*, no justamente una visión general, que incluye la generalización de conceptos en el ámbito del conocimiento y la prácti-

ca. Desde esta perspectiva, la investigación evaluativa es mucho más que la evaluación de programas e incide en procesos, instituciones y otros muchos más objetos.

El cuarto elemento distintivo de la visión transdisciplinar de la investigación evaluativa es que se trata de una visión *técnica*. La evaluación no solo necesita el apoyo técnico de otras muchas disciplinas, sino que, además, tiene su propia metodología. La lógica de la síntesis de resultados, las consecuencias, etc., y la correcta ubicación en el proceso de muchas técnicas auxiliares en las que, probablemente, no es necesario ser un gran especialista, pero sí tener un conocimiento cabal.

Esta *perspectiva transdisciplinar* de la investigación evaluativa de Scriven (1994) coincide en gran medida con los planteamientos que de la misma hemos defendido en otros momentos (Escudero, 1996), pero nosotros no tenemos unas posiciones contrarias a las otras visiones en la misma medida que las tiene Scriven, y, de hecho, consideramos, desde una *posición pragmática*, que todas las visiones presentan puntos fuertes y que, en todo caso, aportan algo útil para la comprensión conceptual y el desarrollo de la investigación evaluativa. Sin embargo, sí que pensamos que esta moderna visión de Scriven es sólida y coherente y ampliamente aceptada en la actualidad.

Una crítica que podría hacerse al planteamiento de Scriven es su excesivo énfasis en la orientación al cliente, al usuario en sentido estricto. Pensamos que esta orientación debe integrarse dentro de una orientación a los implicados, donde existen distintos tipos y distintas audiencias, y, por supuesto, una muy importante son los usuarios en el sentido de Scriven, pero nos parece que la investigación evaluativa hoy en día debe tener una *orientación prioritaria* más plural que la defendida por este autor.

En la última década tenemos una corriente bastante activa de evaluadores que se adscriben a una visión denominada *evaluación basada en la teoría* y que defienden, frente a las posiciones de Scriven y Stufflebeam, más laxas en este sentido, la necesidad del evaluador de contar previamente con un modelo lógico o teórico del programa a evaluar (Shadish, 1998; Christie, 2003; Donaldson y Scriven, 2003; Mark, 2003; Weiss, 2004; Donaldson y Lipsey, 2006; Donaldson, 2007).

Esta visión recoge planteamientos diversos, pero defiende la contextualización y fundamentación específica de cada investigación, de acuerdo con las características particulares del problema, programa u objeto a evaluar, de los implicados en el mismo y de las posibles vías de acción más eficaces. La adaptabilidad, la flexibilidad, la integración de enfoques y el eclecticismo metodológico son elementos definitorios de esta aproximación de la investigación evaluativa, que parte del desarrollo de una teoría del programa, desde distintas fuentes y procedimientos, desde donde plantear las cuestiones de evaluación prioritarias a responder.



## EL MARCO INVESTIGADOR DEL EVALUADOR

En las páginas anteriores hemos recalcado que la investigación evaluativa es un tipo de investigación aplicada que tiene unos objetivos y un terreno de maniobra particular, que no son los que tiene la investigación en general. De hecho, ya hemos visto que incluso se ha venido buscando afanosamente por los epistemólogos y teóricos de la investigación un enfoque paradigmático diferenciado en el que se incluyera de manera coherente la investigación evaluativa.

Nosotros estamos absolutamente de acuerdo con Rea-Dickins (1994) cuando dice que no ve que existan, que operen, linderos o límites diferentes en la investigación evaluativa y en la investigación más convencional. Entre las dos se da un amplio espacio de aspectos comunes, a menudo presentes y solapados en múltiples y diferentes combinaciones. Está muy extendida entre los teóricos la idea de que es el objetivo pretendido por el trabajo, más que la naturaleza del mismo, lo que distingue a la investigación evaluativa de otros tipos de investigación (Clarke y Dawson, 1999).

Sin duda, investigación y evaluación, en el sentido de investigación evaluativa, tienen muchísimo en común y, desde luego, tal como señala Norris (1993), *cuando se*

*habla de investigadores y evaluadores no se está hablando de dos comunidades científicas diferentes.* La relación entre los dos se concibe a menudo como un continuo. Por ejemplo, en el ámbito metodológico ambos comparten raíces, métodos, técnicas y destrezas (Nisbet, 1980; MacDonald, 1987). Sin embargo, no es menos cierto que existen unas peculiaridades que definen *un marco investigador específico del investigador evaluativo*, que fue definido por primera vez como tal, en comparación con el del investigador básico, por Glass y Worthen (1971), y que desde entonces es un tema comentado y debatido de manera regular en artículos y textos sobre investigación evaluativa en general y en algunas de sus diferentes áreas temáticas o aspectos (Anderson y Ball, 1983; Isaac y Michael, 1984; Anderson, 1990; Norris, 1993; Rea-Dickins, 1994; Tejedor y otros, 1994; Escudero, 2000). Algunos de los elementos peculiares de la *investigación evaluativa* frente a la *investigación básica* los ofrecemos a continuación, pero dejando claro que presentamos las situaciones extremas de un continuo, a fin de plasmar de manera más evidente las diferencias. Estas son las siguientes:

#### A. *Motivación*

- a) La investigación básica persigue satisfacer una curiosidad. Fundamentalmente, el investigador está *intrigado*.
- b) La investigación evaluativa persigue contribuir a la solución de un problema particular. El evaluador, y desde luego su promotor, está *preocupado*.

#### B. *Objetivos, finalidad o propósito*

- a) La investigación básica busca *conclusiones* que incrementen el cuerpo teórico del conocimiento sobre el tema.

- b) La investigación evaluativa busca ofrecer sugerencias para tomar *decisiones* al servicio de la resolución de un problema, de un plan, etc. (Tukey, 1960; Cronbach y Suppes, 1969).

C. *El punto de partida*

- a) La investigación básica parte de la formulación de hipótesis que hay que «someter a prueba», aceptándolas o rechazándolas.
- b) En la investigación evaluativa resulta difícil e inadecuado formular hipótesis precisas (De la Orden, 1985) porque el propósito de la evaluación es mejorar, no probar (Stufflebeam y otros, 1970). Se suele partir de varias opciones alternativas a la vez, y no solo de una.

D. *Productos, resultados*

- a) La investigación básica es fundamentalmente *nomotética* (dar leyes). Busca el estudio y establecimiento de leyes, de relaciones entre variables.
- b) La investigación evaluativa es fundamentalmente *ideográfica* (descriptiva del caso particular). Trata de describir algo en particular, en relación con una o varias escalas de valor.

E. *El papel de la explicación*

- a) La explicación es fundamental en la investigación básica.
- b) Una buena investigación evaluativa puede no necesitar explicar por qué un programa o un sistema tiene buenos rendimientos. Puede bastarle con constatarla.

F. *El papel del investigador (autonomía)*

- a) En la investigación básica es el propio investigador el que selecciona y establece el problema o

cuestiones que desea investigar. La investigación básica es, en principio, una *actividad independiente y autónoma* que establece sus propios problemas y actividad.

- b) En la investigación evaluativa se responde generalmente a una demanda. *El evaluador responde a la solicitud* de un «cliente». Aquí surge el problema del carácter del auditor-juez que el evaluador podría tener para implicados en el programa, produciendo relaciones y situaciones problemáticas, no presentes en otros tipos de investigaciones (MacDonald, 1976; Weiss, 1975; House, 1989).

#### G. *Propiedades del fenómeno estudiado*

- a) La investigación básica puede obtener una evidencia de la utilidad social de lo que estudia.
- b) La investigación evaluativa trata de analizar directamente la *utilidad social* del fenómeno estudiado.

#### H. *Generalidad o universalidad del fenómeno estudiado*

- a) La generalidad de los resultados es un fin primordial en la investigación básica.
- b) La investigación evaluativa se centra en un fenómeno cuya generalización se limita al *contexto envolvente* del mismo. Sin embargo, hallazgos similares en contextos diversos pueden desembocar en nuevo conocimiento generalizable.

#### I. *Importancia del interrogante de valor*

- a) En la investigación básica, averiguar el valor de los casos no es un objeto directo.
- b) En la investigación evaluativa las preguntas de valor son la condición sine qua non y suelen determinar la información que se busca.

## J. *Métodos de investigación*

- a) El investigador básico está mucho más preocupado por el control de variables para certificar y generalizar sus reglas y explicaciones causales.
- b) El investigador evaluativo, aun apoyándose en una metodología similar, está menos preocupado por el citado control porque no le interesan tanto las relaciones causales cuanto las concomitancias de un programa, un plan, etc., con unos determinados efectos.

## K. *Criterios de validez*

- a) La investigación básica requiere *validez interna* (los resultados son inequívocos y no existe confusión de variables ni errores sistemáticos) y *validez externa* (los resultados pueden generalizarse a otros contextos).
- b) La investigación evaluativa necesita *isoformismo* entre la información recogida y la realidad, *credibilidad* ante los clientes implicados, *utilidad* y *factibilidad* (Patton, 1984).

## L. *Base disciplinar*

- a) El investigador básico define libremente sus propios problemas y rara vez se hace preguntas que le lleven fuera de la *disciplina* en la que está especializado.
- b) El investigador evaluativo goza de menos libertad para establecer las preguntas que debe contestar, necesitando a menudo abarcar un *rango disciplinar más amplio* e incluso *multidisciplinar*.

## M. *Tiempo de realización*

- a) La investigación básica suele realizarse *sin límites temporales*.

- b) La investigación evaluativa suele tener *plazos de realización*.

#### N. *Técnicas instrumentales*

- a) En la investigación básica las técnicas instrumentales están muy *condicionadas* por el control de la validez interna y externa.
- b) En la investigación evaluativa dichas técnicas son mucho más flexibles y diversificadas, para evitar la pérdida de cualquier información relevante.

#### Ñ. *Audiencia*

- a) La investigación básica se dirige a una *audiencia universal*, solamente condicionada por su interés y preparación.
- b) La investigación evaluativa se dirige fundamentalmente a las *audiencias implicadas*, o afectadas, en el objeto evaluado (*stakeholders*).

#### O. *Implicación de valores*

- a) La investigación básica, aunque no está libre de valores, se plantea con *vocación de independencia*.
- b) La investigación evaluativa tiene el riesgo de estar mucho más condicionada por los valores de los que la realizan, pues se juzga la eficacia, eficiencia, etc., de proyectos sociales, en un *contexto social contaminado de valores* (Weiss, 1975; Anderson y Ball, 1983).

#### P. *Presión externa*

- a) La investigación básica se suele llevar a cabo en un contexto sin presión o con *escasa presión externa*. Frecuentemente, se lleva a cabo en el anonimato.
- b) La investigación evaluativa se suele llevar a cabo en un *contexto de presión externa*, sociopolítica, a la espera de sus resultados y deseando que sean favorables

para las diversas audiencias (House, 1973; Escudero, 1980).

#### Q. *Informe de la investigación*

- a) Los informes de la investigación básica se elaboran según unas normas convencionales de cada disciplina y se difunden fundamentalmente en *revistas o libros especializados*.
- b) La investigación evaluativa suele concluir en *informes* (memorias) *amplios*, orientados para las distintas audiencias implicadas.

#### R. *Vigencia de los resultados*

- a) Dicha vigencia puede ser *larga*, y en muchos casos lo es, en la investigación básica.
- b) En la investigación evaluativa casi siempre es *corta* y está muy vinculada al contexto concreto del momento del estudio.

Anderson (1990) compara la *investigación académica* y la *investigación por contrato*, en el supuesto de que mucha de la investigación evaluativa es, de hecho, investigación por contrato. De este análisis podríamos exponer algunos detalles que no se muestran de manera explícita en los puntos anteriores:

- a) La investigación evaluativa se suele *contratar*, la académica *financiar*.
- b) La investigación académica se apoya más en la teoría, literatura e investigación previas que la investigación evaluativa, más centrada en el caso concreto.
- c) La investigación evaluativa tiene una aplicación más obvia y directa que la investigación académica.
- d) Los resultados de la investigación por contrato pertenecen al que contrata, los de la investigación académica pertenecen al investigador.

A pesar de reconocer estas diferencias, estamos absolutamente de acuerdo con Gephart y Potter (1976) en que se trata de simples variantes del mismo método científico, y con Weiss (1972) y De la Orden (1985) en que la diferencia esencial es más de orientación que de método. Incluso existen autores (Anderson y Ball, 1983; Rutman, 1984) que defienden el potencial valor de la investigación evaluativa en la producción de resultados generalizables, sobre todo compaginando resultados en instancias diversas (Cronbach, 1982).

La metaevaluación, *la evaluación de la propia evaluación*, introducida por Scriven (1969), adquiere importancia como lógica consecuencia del desarrollo de la investigación evaluativa y de la expansión de proyectos de evaluación en todos los terrenos. Parece razonable pensar que lo primero que se necesita es *confianza* en algo que, precisamente, se hace para *tener confianza* a la hora de tomar decisiones, intervenir o sugerir algo sobre un proyecto social. Se trata de articular la obligación científica de especular sobre lo que puede estar yendo bien o mal con un plan determinado (Escudero, Pino y Rodríguez, 2010).

Un autor de larga tradición en el tratamiento del tema como Stuffebeam (2001), además de indicar que la metaevaluación debe juzgar el mérito, el valor y la utilidad de la evaluación, la establece como elemento de garantía científico-profesional; los propios evaluadores deben asegurar que las investigaciones evaluativas estén bien hechas, que se mejoren sistemáticamente y que den garantía a sus audiencias. Por ello, la metaevaluación es una obligación profesional de los evaluadores, que entronca con la ética de la revisión y el autocuestionamiento continuos (Stake, 2006). La metaevaluación in-

tenta detectar sesgos de tipo estadístico, desviaciones en la interpretación de los datos, problemas de comprensión de algunos aspectos, etc., dentro de una perspectiva crítica que parte de la idea de que toda evaluación es mejorable (Straw y Cook, 1985).

Siguiendo al propio Stufflebeam (2001), podemos definir *metaevaluación* como «el proceso de diseño, obtención y aplicación de información descriptiva y valorativa —acerca de la utilidad, factibilidad, legitimidad y rigor de una evaluación, así como de su sistematicidad y de la competencia, integridad, honestidad, respetabilidad y responsabilidad social de sus autores— sobre el desarrollo de la evaluación que informa públicamente de sus fortalezas y debilidades» (185). Este mismo autor nos dice que un buen símil para entender este concepto es que la *metaevaluación* es en el campo de la *evaluación* lo que la *auditoría* en el campo de la *contabilidad*, que asegura el escrutinio externo e independiente del uso razonable de los recursos financieros.

En la metaevaluación también confluyen las perspectivas *formativa* y *sumativa*. En la primera, como una guía de ayuda a los evaluadores a preparar todo antes de realizar la evaluación; en la metaevaluación sumativa se determina el mérito, valor y utilidad de la evaluación una vez terminada. Ambas perspectivas determinan conjuntamente el grado en que la evaluación es técnicamente adecuada, útil para la toma de decisiones, ética en su incidencia sobre personas y eficiente en el uso de recursos. En la aproximación formativa el metaevaluador actúa como un *técnico consultor* y en la sumativa como un *agente externo certificador*.

El trabajo metaevaluador puede llevarse a cabo según distintos enfoques modélicos. Los más habituales son los siguientes (Posavac y Carey, 1989): *a*) revisión del estu-

dio, b) análisis secundario y c) integración cuantitativa. La *revisión del estudio* se basa fundamentalmente en el análisis del informe de evaluación, de los procesos y resultados; es un procedimiento esencial en el estudio de naturaleza interpretativa, en especial sobre datos cualitativos. El *análisis secundario* supone el reanálisis de los datos originales, siendo un procedimiento muy poderoso si se puede llevar a cabo correctamente. El hecho de disponer los datos de una evaluación ya es signo de credibilidad de la misma. La *integración cuantitativa* o meta-análisis (Glass y otros, 1981) es un procedimiento estadístico de combinación de resultados cuando existen varios estudios sobre el mismo tema. También podemos hablar de metaevaluaciones internas y externas.

Las metaevaluaciones se apoyan en técnicas muy variadas, aunque existen algunas de creciente utilización, como los paneles de revisión, análisis de protocolos, seminarios interactivos y las listas de verificación (*check-lists*) adecuadas para evaluaciones de distinta naturaleza y contenido (Scriven, 2000; Stufflebeam, 2000).

Para juzgar la calidad de las evaluaciones se han ofrecido diversos criterios, en general ligados con el mérito y el valor, la utilidad, la eficiencia del diseño, etc. (Stufflebeam y otros, 1970; Joint Committee, 1981; Schwandt y Halpern, 1988). Así, se habla de criterios de a) *rigor* (métodos precisos correctamente aplicados), b) *utilidad* (valor, mérito, importancia de resultados) y c) *eficiencia y efectividad* del diseño y del proceso evaluador. También se habla de criterios *éticos*, normalmente como protección de los derechos de los implicados en la evaluación, preservando la verdad de lo evaluado, en respuesta a determinadas presiones distorsionadoras.

En relación con el desarrollo de la metaevaluación, en un contexto de crecimiento profesional de la investiga-

ción evaluativa, está el establecimiento de *normas* bajo las que desarrollar la evaluación con vistas a asegurar su calidad y evitar el máximo de potenciales defectos. Durante las últimas décadas se han elaborado normas sobre investigación y medición educativas por distintas asociaciones científicas, tales como la American Psychological Association (APA), la American Educational Research Association (AERA) y el National Council on Measurement in Education (NCME).

Pero a nadie se le escapa que el prestigio académico y social que la investigación evaluativa como disciplina y la tarea de los evaluadores profesionales han venido adquiriendo y aumentando en las últimas décadas está en buena medida marcado por el liderazgo y tutela del Joint Committee on Standards for Educational Evaluation, creado por las tres asociaciones científicas antes citadas, que empezó a trabajar en 1975, dos décadas después de que Cronbach impulsara desde la American Psychological Association (1954) el desarrollo de estándares para ayudar a la sociedad a protegerse de las prácticas corruptas de las compañías que desarrollaban y comercializaban tests.

Este Comité está actualmente auspiciado por más de una quincena de organizaciones académico-profesionales, fundamentalmente americanas y canadienses, del ámbito de la evaluación educativa (Sanders, 1994; Joint Committee, 2003), y su mayor aportación ha sido la elaboración y el establecimiento de unas *normas ético-científicas*, de unos estándares, para regular la evaluación educativa en Estados Unidos y Canadá. Sin embargo, su aplicación ha ido mucho más allá del aludido espacio geográfico y del ámbito de la educación, aunque se hayan suscitado algunos problemas de adaptación en determinados contextos (Stufflebeam, 2004).

Las primeras normas para evaluación de programas educativos fueron publicadas en 1981, en 1988 se publicaron las normas para evaluación de educadores y en el 2003 las de estudiantes (Joint Committee, 1981, 1988 y 2003). En el año 1989, estas normas fueron certificadas por el American National Standards Institute (ANSI), convirtiéndose de hecho en unas normas oficiales en los Estados Unidos para la evaluación de programas. En 1994 el Comité revisó las normas de evaluación de programas y, una vez publicadas, se convirtieron en las primeras normas aprobadas por el ANSI sobre principios y conductas profesionales (Joint Committee, 1994).

El objetivo básico del Comité Conjunto y de sus normas, que se entienden como algo dinámico y que debe revisarse y adecuarse periódicamente, es promover la concienciación sobre la necesidad de que las evaluaciones sean sólidas, de calidad, útiles y que respondan a las necesidades de evaluación. Para conseguir esto, el Comité establece normas de los cuatro tipos siguientes: de *utilidad*, de *factibilidad*, de *legitimidad* y de *precisión*.

Es muy importante observar el orden en el que se plantean las distintas normas, pues actúan como llave o condición previa para las siguientes. En concreto, al plantearse una evaluación, la primera condición que hay que asegurar es que sea *útil*, esto es, que tenga sentido llevarla a cabo para mejorar el sistema, tomar decisiones sumativas, planificar, etc. En otras palabras, el Comité Conjunto nos viene a sugerir, en alguna medida, que, de entrada, una evaluación tiene que ser *útil*, y que, si esto no es así, no tiene sentido plantearla.

Asegurada la utilidad de una evaluación, es cuando hay que analizar si es *viable*, si se puede hacer; después, si es *legítima*, si no viola principios de legitimidad, derechos de implicados, etc.; y, finalmente, se analiza,

diseña y se realiza con el máximo *rigor científico*, siguiendo los principios y la metodología más adecuados y poderosos.

Hoy en día, las bases para este proceso metaevaluativo —no olvidando que el contexto natural de su definición son los Estados Unidos y Canadá, con lo que su aplicación en otros contextos requiere contextualización (Jang, 2000)— están en los estándares de evaluación de personal y de programas educativos del Joint Committee (1988 y 1994) y en los principios guía de la American Evaluation Association (1995), que, aunque no son tan precisos y detallados como los anteriores, están definidos para todos los ámbitos, no solo el educativo, y, además, se ha demostrado (Sanders, 1995; Grasso, 1999) que son compatibles y complementarios.

Parece indudable que la gran importancia e influencia que la investigación evaluativa ha tenido en las últimas décadas en la educación y en otros ámbitos sociales habría sido más cuestionada si las asociaciones profesionales más directamente implicadas no hubieran establecido estas *normas ético-científicas* para regular el trabajo de los evaluadores. Un resumen muy sucinto de estas normas es el siguiente:

#### A. *Normas de utilidad*

Las normas de utilidad tienen el propósito de asegurar que una evaluación satisface las necesidades prácticas de información de las audiencias. Lo que conlleva:

- A1 Identificación de las audiencias para satisfacer sus necesidades.
- A2 Confiabilidad, competencia y credibilidad del evaluador.
- A3 Selección y alcance de información pertinente para responder a necesidades e intereses.

- A4 Interpretación valorativa de resultados clara y fundamentada.
- A5 Claridad del informe para su comprensión por las distintas audiencias.
- A6 Difusión del informe a todos los implicados e interesados.
- A7 Oportunidad del informe para poder utilizarse a tiempo.
- A8 Trascendencia de la evaluación para estimular su utilización.

### B. *Normas de factibilidad*

Las normas de factibilidad tienen el propósito de asegurar que una evaluación es realista, prudente, diplomática y frugal. Lo que conlleva:

- B1 Procedimientos prácticos con la información necesaria.
- B2 Viabilidad política.
- B3 Relación costo-producto adecuada.

### C. *Normas de legitimidad*

Las normas de legitimidad tienen el propósito de asegurar que una evaluación se dirige, legalmente y con ética, en beneficio de quienes participan en la evaluación y de aquellos a quienes los resultados les afectan. Lo que conlleva:

- C1 Seguimiento de las obligaciones formales acordadas.
- C2 Gestión equilibrada y honesta de los conflictos de intereses.
- C3 Exposición total y franca de resultados y limitaciones del estudio.
- C4 Respeto a los derechos a la información pública.
- C5 Respeto a los derechos individuales.

- C6 Respeto a valores sociales y colectivos.
- C7 Equilibrio del informe al presentar aciertos y defectos del objeto evaluado.
- C8 Responsabilidad fiscal en la distribución y asignación de recursos.

#### D. *Normas de precisión*

Estas normas tienen el propósito de asegurar que una evaluación revela y proporciona la información técnicamente apropiada acerca de los rasgos del objeto de estudio que determinan su valor o mérito. Lo que conlleva:

- D1 Identificación clara del objeto a evaluar.
- D2 Análisis detallado del contexto y de sus posibles influencias.
- D3 Propósitos y procedimientos claramente descritos.
- D4 Fuentes de información identificadas y confiables.
- D5 Medición válida.
- D6 Medición confiable.
- D7 Control sistemático de datos.
- D8 Análisis apropiado de la información cuantitativa.
- D9 Análisis apropiado de la información cualitativa.
- D10 Conclusiones fundamentadas.
- D11 Informe objetivo, libre de distorsiones y sesgos.

La revisión de las normas de 1994 introdujo algunos cambios con relación a las de 1981, siendo, según Rogers (1994), la *nueva visión de validez en la medición* el más importante de todos ellos. Se destaca que son las inferencias que se hacen a partir de la evaluación lo que debe ser validado, más que la propia recogida de la información. La palabra *inferencia* incluye interpretación y uso. También se resalta la multiplicidad de fuentes de infor-

mación para validar las decisiones, porque las fortalezas de unos procedimientos compensan las debilidades de otros (Joint Committee, 1994).

Otro elemento renovador, y saludable según LeCompte (1994), es la apertura a las metodologías de corte cualitativo en la evaluación y a la participación de todos los implicados y afectados por ella. También se hace eco LeCompte de la nueva concepción de la validez y de la ampliación de perspectiva en cuanto a procedimientos de recogida y análisis de información. Las nuevas normas, en suma, legitiman la integración en la evaluación de todo el amplio rango de procedimientos que nos ofrece la metodología de la investigación.

Thompson (1994) señala, a este respecto, que parece que las nuevas normas lo que hacen es incorporar lo que la literatura especializada reconoce como buenas prácticas evaluadoras, al tiempo que se hace eco de estrategias de actuación de los evaluadores y de recomendaciones para mejorar el nivel de utilización y *optimizar el uso de las evaluaciones*. Estas recomendaciones están en la línea de *a)* identificar el motivo de la evaluación, *b)* concentrar esfuerzos sobre las necesidades prioritarias de las distintas audiencias implicadas, *c)* reconocer la subjetividad de la evaluación, *d)* considerar su realidad política, *e)* recomendar explícitamente decisiones de política, *f)* no resaltar demasiado fórmulas aisladas de prueba y *g)* construir una relación positiva con los administradores y el personal del programa.



## UTILIDAD Y USO DE LA EVALUACIÓN

El análisis y las reflexiones anteriores nos confirman que, en el momento presente, en pocas cosas están tan de acuerdo los expertos en investigación evaluativa como en el hecho de que esta se lleva a cabo con la intención de ser útil e influir en la toma de decisiones posterior. De hecho, los propios evaluadores dedican grandes esfuerzos a conseguir que sus sugerencias sean tenidas en cuenta y aplicadas convenientemente. Sin embargo, la utilización de los resultados de la evaluación, por múltiples razones, siempre es un asunto controvertido.

Un elemento crítico en este debate es el hecho de que «la evaluación es una acción política en un contexto en el que el poder, la ideología y los intereses son soberanos e influyen en las decisiones más que las sugerencias de la evaluación» (Shadish, Cook y Leviton, 1991: 448). Por ello, es muy pertinente la observación de Patton (1997), pensando en los evaluadores, cuando dice que «ser inocente ante la naturaleza política de la evaluación supone convertirse, consciente o inconscientemente, en un peón del juego de alguien» (291).

La utilización de la evaluación, siguiendo a Weiss (1986), va a estar condicionada por la perspectiva que toma el propio proceso evaluador, esto es, por la orienta-

ción principal que sustenta la investigación evaluativa que se está llevando a cabo. En este sentido, se habla de siete orientaciones:

- *Modelo de desarrollo del conocimiento*: orientación derivada de la investigación aplicada en general.
- *Modelo de resolución de problemas*: se busca la clase de información que se necesita para tomar decisiones y resolver un problema planteado.
- *Modelo interactivo*: la hipótesis de partida es que la información para la toma de decisiones no solo la proporciona la evaluación, sino también el trabajo interactivo con políticos, administradores, clientes y otros grupos diversos.
- *Modelo político*: en ciertas circunstancias, las posiciones políticas de quienes toman las decisiones pueden hacerlos insensibles a ciertos resultados de la evaluación. Solo les interesará lo que refuerce sus posiciones.
- *Modelo táctico*: el contenido y los resultados de la evaluación importan poco; es solo el hecho de que se está haciendo lo que se utiliza para evitar críticas o demorar decisiones.
- *Modelo de la ilustración o iluminación*: las evaluaciones son el mejor camino de informar al gran público del estado de las cosas y, así, influir en las políticas sociales.
- *Modelo de la empresa intelectual*: la evaluación se entiende como un elemento más en la empresa intelectual de la sociedad y en la configuración del pensamiento social.

Probablemente, no existe ningún ejemplo de investigación evaluativa que siga una de estas orientaciones de manera exclusiva, pero sí es cierto que, aunque la mayoría se apoyan en varias de ellas, una suele ser preponde-

rante en cada caso, condicionando de manera decisiva el campo y el alcance de la utilización de sus resultados.

Es la propia Weiss (1998) la que nos dice que podemos distinguir cuatro usos de la investigación evaluativa, que son bastante diferentes en su naturaleza y que tienen conexión con las orientaciones o modelos antes citados.

En primer lugar, podemos distinguir el *uso instrumental*, que está relacionado con la toma de decisiones sobre el objeto evaluado, y que suele tener lugar cuando se dan tres condiciones: 1) las implicaciones de los resultados no son controvertidas, ni provocan grietas en la organización, ni conducen hacia un conflicto de intereses; 2) los cambios demandados están dentro del repertorio previsto por el programa y son de poco alcance; y 3) el contexto del programa se mantiene relativamente estable, sin grandes cambios en el liderazgo, presupuestos, tipos de clientes y apoyo social. Hay una cuarta situación en la que el uso se potencia también, y es cuando un programa está en crisis o en situación de parálisis y nadie sabe qué hacer. En ese caso, como situación extrema, se echa mano de la evaluación para intentar resolver el problema.

En segundo lugar, se distingue el *uso conceptual* de la evaluación, que sirve para cambiar de manera progresiva, entre los responsables, las concepciones de los programas, de lo que son y de lo que hacen. Cuando las condiciones organizativas se tornan favorables, los nuevos planteamientos pueden conducir a un uso instrumental del primer tipo.

Un tercer tipo de uso es el que podría denominarse de *apoyo o persuasión*, pues la evaluación sirve para conseguir apoyo para cambios necesarios y demandados en los programas. Se usa la evaluación como elemento de persuasión para ganar adeptos en defensa del cambio.

El último tipo de uso es el que *transciende el objeto evaluado*, pues acumula conocimiento en otras muchas evaluaciones, para contribuir a cambios en el pensamiento y en las políticas a gran escala. Este no suele ser un objetivo directo de los evaluadores, pero es una categoría de utilización que no debe ser olvidada y que entronca con el concepto de *influencia*, que comentamos posteriormente.

En todo caso, existen medios para potenciar la utilización de las evaluaciones. Alkin (1991) sugiere algunos, como identificar a los responsables clave en la toma de decisiones, asegurarse de que la evaluación plantea las cuestiones formuladas por los implicados, limitar el estudio a las variables bajo el control del que toma las decisiones y comunicar los resultados lo antes posible, con informes cortos, concisos y claros. Leviton (2003) viene a señalar algo parecido cuando aconseja focalizar bien los contextos de decisión de los usuarios potenciales, porque es necesario ajustar los resultados de la evaluación a las construcciones de la realidad de los usuarios y a la experiencia que aportan al propio contexto.

Por razones similares a las anteriores, Patton (1997) defiende estrategias participativas en la investigación evaluativa, esto es, la colaboración con los implicados, como medio de potenciar la utilización de los resultados. Algo que él viene a denominar *el uso intencionado de los usuarios previstos*. Sin embargo, no podemos obviar el riesgo de conservadurismo que tiene la participación de los responsables en la evaluación (Weiss, 1998), pues difícilmente van a querer que se cuestionen los principios fundamentales y el statu quo de la organización o de los programas evaluados.

Rossi y Freeman (1993) también señalan algunas reglas a seguir por los investigadores, con vistas a maximizar

zar la utilización de sus evaluaciones: *a)* conocer y comprender cómo operan los que toman las decisiones, *b)* los resultados de la evaluación deben estar en el momento en el que son necesitados, *c)* la evaluación debe respetar los compromisos de los implicados en los programas, *d)* los planes de utilización y diseminación deben ser parte del diseño de la evaluación y *e)* las evaluaciones deben incluir una evaluación de su utilización.

Reflexionando desde la práctica, Grasso (2003) nos dice que, de antemano, nunca vamos a tener completa seguridad de que una evaluación se utilizará, pero que esta posibilidad se acrecienta si se optimizan las condiciones para ello. En concreto, una evaluación es más útil y más *utilizable* si el evaluador ha respondido previamente las tres preguntas siguientes:

- ¿Quiénes utilizarán la evaluación?
- ¿Qué necesitarán de la evaluación?
- ¿Cuándo necesitarán la información?

La primera pregunta busca determinar cuáles son las *audiencias*, cómo se les puede hacer llegar la información, cuáles son las prioridades de cada una de ellas y con qué prioridad se les debe atender. Por todo ello, es fundamental el contacto y las consultas con las diversas audiencias antes de diseñar y llevar a cabo la evaluación.

La segunda pregunta incide directamente en la comunicación comprensible de los resultados de la evaluación, no solamente intentando responder a las necesidades de las audiencias, sino fundamentando lo mejor posible y de la manera más clara las sugerencias de acción. En muchos casos, los ejemplos y las recomendaciones concisas suelen ser más útiles que las explicaciones más completas y documentadas.

En tercer lugar, hay que resaltar que el tiempo es crítico para que los resultados de la evaluación tengan la

consideración adecuada. No se trata tanto de evaluar de forma rápida, sino de proporcionar la información y las sugerencias en el momento preciso. Siempre con la conciencia de que los resultados de la evaluación no van a ser el único criterio válido para tomar decisiones, pero sí uno de los importantes. También hay que aprovechar el potencial de la evaluación para ofrecer información no buscada y no esperada (*serendipity*), que puede ser de gran utilidad en la resolución de diversos problemas.

### **9.1. Comunicar y diseminar la información adecuadamente**

Es importante señalar que partiendo de principios como los descritos se desarrollan y aplican distintos sistemas de auditoría de la información (SAI) producida por las evaluaciones (Schwartz y Mayne, 2005) para asegurar su calidad y, al mismo tiempo, que esté disponible cuando sea preciso, que responda a las demandas del problema y que sea comprensible para los que la necesitan y demandan.

Parece muy claro que una buena diseminación de los resultados es clave para incrementar la probabilidad de utilización de una evaluación (Clarke y Dawson, 1999), pero el informe final no es el único medio para hacer llegar la información a los implicados. Las presentaciones formales y las comunicaciones informales son procedimientos de obligado uso para informar sobre resultados de las evaluaciones.

Los informes de evaluación no deben confundirse, en su estilo, con artículos científicos, pues deben ser comprendidos por implicados con formación académica muy diversa. Además, deben contener un apartado sumario de los resultados y sugerencias relevantes para

la toma de decisiones. A veces se habla (Rossi y Freeman, 1993) de una *diseminación primaria*, para identificar el informe final completo y técnico de la evaluación, y de *diseminación secundaria*, al referirse a informes concretos, orales y escritos, dirigidos a diversos sectores implicados.

Owen y Rogers (1999) ofrecen un análisis del espacio que va desde los *resultados de la evaluación* hasta su *utilización*. Nos hablan de una secuencia que tiene tres grandes pasos:

- *Diseminación* o el conjunto de estrategias deliberadas usadas por los evaluadores para informar a las audiencias identificadas acerca de la evaluación, de sus resultados y de las implicaciones para las mismas. La diseminación es más dependiente de las visiones y conductas de los evaluadores que de las características y disposición al cambio de las audiencias implicadas.
- *Ilustración, iluminación, esclarecimiento* o uso conceptual de la evaluación (Alkin, 1991), que significa el esfuerzo hecho por los evaluadores para responder a las necesidades de las distintas audiencias implicadas a fin de que planifiquen su actuación y tomen decisiones con conocimiento de causa sobre el programa u objeto evaluado. Este uso depende de las estrategias del evaluador para responder a posibles necesidades de las audiencias, de la comunicación con ellas y de la disposición al cambio de las mismas.
- *Otros usos*, fundamentalmente los instrumentales, que dependen mucho más de la actitud y disposición de las audiencias que de las estrategias de los propios evaluadores. En este apartado, se suele hablar de tres tipos de usos:

- *Uso de legitimación*, esto es, para *justificar* decisiones previamente tomadas, apoyándose en otros planteamientos. Es una evaluación retrospectiva, de verificación, de apoyo a decisiones previas.
- *Uso simbólico*, esto es, cuando la evaluación se realiza como pretexto justificativo de algo, o para favorecer otro proceso, pero sin intención de utilización real. Por ejemplo, a efectos de imagen, puede ser de interés, para una institución política u otro tipo de organización, tener en marcha un programa de evaluación al que, realmente, no se le hace caso.
- *Uso instrumental*, esto es, el uso que se hace de los resultados de la evaluación para la acción. Las decisiones, formativas y sumativas, pueden ser de muy diverso tipo y afectar a todo tipo de implicados y elementos del objeto de la evaluación.

También existe una corriente de evaluadores (Kirkhart, 2000; Henry y Mark, 2003) que defienden que el concepto de uso de la evaluación, entendido como utilización directa y focalizada, se ha quedado estrecho, por lo que debe dejar paso al concepto de *influencia*, como idea de consecuencias más abiertas y amplias, y no necesariamente delimitadas de antemano. La influencia habla también de las consecuencias posibles, y no solo de las buscadas, esto es, de la capacidad de producir cambios por medios intangibles o indirectos.

Profundizando en la anterior línea de pensamiento, Henry (2003) enfatiza el gran valor de la evaluación en las democracias, como instrumento de mejora social, aunque, en su opinión, los evaluadores no han sabido explotar todo su potencial para influir en este sentido, probablemente porque la investigación sobre el uso de la evaluación se ha centrado en su utilización inmediata.

En la práctica, algo que condiciona extraordinariamente el uso o influencia de la evaluación es la manera en que se comunica o se informa de ella. Es importante resaltar que la buena investigación evaluativa no solo se fundamenta en buenas técnicas analíticas, como cualquier otro tipo de investigación social, pues debe combinar aquellas con técnicas de comunicación bien desarrolladas. En otras palabras, no basta con que la investigación evaluativa esté bien hecha; además, debe estar bien contada, no solo a los especialistas, sino a todos los implicados y afectados por la misma, a todas las audiencias. A esto hay que añadir que hay especialistas (Perrin, 2001) que ponen mucho énfasis en la conveniencia de que los evaluadores se apoyen siempre en los métodos más simples de entre los que permitan responder a las preguntas planteadas, pues esto no solo tiene consecuencias en los costes, rapidez y aceptación de la evaluación, sino también en su uso.

Por lo anterior, existen distintos planteamientos sobre cómo se debe informar de las evaluaciones con vistas a *optimizar su visibilidad* y, consecuentemente, favorecer su utilización. A partir de un informe común del proceso de evaluación y sus resultados, se pueden introducir subapartados, apéndices, sugerencias específicas, etc., dirigidos a las distintas audiencias implicadas. Otra opción es elaborar informes separados según las audiencias, aunque esta es una fórmula menos habitual.

Lo que sí resulta fundamental, en todo caso, es la adecuación del lenguaje técnico a la hora de presentar los resultados, que, en la investigación evaluativa, las más de las veces son producto de algún tipo de tratamiento estadístico, que no es familiar para muchos de los potenciales receptores de tales resultados. El evaluador debe ser capaz de presentar sus resultados apoyándose en la jerga

estadística convencional y, además, de hacerlo de forma que tenga significado (May, 2004) para las distintas audiencias.

Este autor nos dice que un resultado estadístico tiene significado para una audiencia determinada cuando es:

- *Comprensible*, esto es, que se entienda de manera fácil por la mayoría de la gente, con el mínimo de supuestos posibles sobre la preparación estadística de la audiencia y evitando la jerga estadística.
- *Interpretable*, esto es, utilizando unidades de medida familiares o fácilmente explicables.
- *Comparable*, esto es, que permita observar las diferencias con otras situaciones, a lo largo del tiempo, otros programas, otros centros, etc.

El tercer criterio, el de la comparabilidad, nos dice May que es menos importante que los dos primeros, pues no siempre se dan las condiciones de necesidad o propicias para la comparación. Cuando esto sucede, el criterio Enriquece mucho el valor de los anteriores.

## **9.2. Estrategias participativas**

Desde hace varias décadas, y por múltiples y variadas razones tanto teóricas como prácticas, los expertos han venido defendiendo la implicación de los afectados (*stakeholders*) en los propios programas de evaluación. Una de esas razones es la de potenciar la utilización de los resultados de la evaluación en los distintos procesos de toma de decisiones que pueden afectar a los diversos implicados. La idea básica es que, si se participa y colabora con la evaluación, es mucho más fácil asumir sus resultados y sugerencias de acción y, consecuentemente, aceptar y promover los cambios necesarios, aunque sean costosos y supongan algunos sacrificios. Con la participa-

ción de los afectados, la evaluación, globalmente, aumenta su viabilidad y utilidad potencial. Además, se refuerza el principio de corresponsabilidad entre los implicados ante los problemas detectados y sus posibles decisiones y se potencian las fórmulas de gestión horizontal en los programas e instituciones.

Una de las claves para que este tipo de participación se produzca es que exista en las instituciones una cultura, esto es, un conjunto de supuestos básicos, normas, valores y visiones culturales compartidos por sus miembros, que influyan en su funcionamiento, integrando lo que podemos denominar *cultura evaluativa* (Scheerens, 2004). En otras palabras, es necesaria la creencia de que la organización aprende y mejora a partir de la reflexión crítica y evaluación sistemática que llevan a cabo sus propios miembros. Una de las maneras más eficaces para vencer muchas de las inhibiciones de, por ejemplo, los profesores ante la evaluación son los procesos de reflexión crítica sobre su propia tarea y el funcionamiento institucional.

Las fórmulas de colaboración y participación en las evaluaciones son múltiples y variadas en cuanto a intensidad e implicación (Taut, 2008), aunque todas ellas se sustentan, de alguna manera, en modelos organizativos reconocidos genéricamente como democráticos. En algunas ocasiones con procesos de una sola fase, en la que todos participan, y en otras con varias fases, con distintos grados de implicación de unos y otros. Una fórmula extraordinariamente extendida en la evaluación institucional, por ejemplo, es conjugar fases de autoevaluación y evaluación externa, buscando la integración de los resultados y la complementariedad de las posiciones e intereses de las distintas audiencias. Algunos autores van todavía más lejos (Christie, Ross y Klein, 2004), pues defienden que ambas evaluaciones sean llevadas a cabo

por un único equipo, compuesto de agentes internos y externos, trabajando en todo momento en colaboración.

Al hablar de la expansión de la evaluación democrática resulta obligado referirse al clásico trabajo de MacDonald (1976), en el que definía tres formas de evaluación: autocrática, burocrática y democrática. En la *evaluación autocrática* el evaluador se posiciona como un experto externo, a cuyo conocimiento se le da un carácter científico, que se usa para garantizar las situaciones y políticas públicas. En la *evaluación burocrática* el evaluador se sitúa dentro del aparato o estructura de poder institucional, fundamentalmente para preservarlo. En la *evaluación democrática*, finalmente, el evaluador actúa como un experto independiente, que media en la distribución de las relaciones de poder en el contexto de la evaluación.

Desde estos planteamientos iniciales de MacDonald, la idea de evaluación democrática se ha ido desarrollando con diversos enfoques y aproximaciones (Ryan, 2004), que se reflejan en los aspectos organizativos y metodológicos, en la generación del conocimiento, en el papel de los evaluadores y en las prácticas de comunicación entre los implicados.

La conocida genéricamente como *evaluación democrática* se fundamenta en la democracia liberal (Locke, 1988), en los valores de pluralismo y ciudadanía informada, se ubica dentro de las instituciones y sistemas formales y se apoya preferentemente en la metodología del estudio de casos. Desde el punto de vista práctico, el evaluador es un gestor, un agente de la información, que atiende a un rango de intereses plural, con negociación, accesibilidad y confidencialidad. Todos los implicados son considerados como una fuente significativa de información para la evaluación y, al mismo tiempo, todos tienen derecho a estar informados.

La *evaluación democrática deliberativa* se diferencia de la anterior en algunos aspectos, como en su fundamento en un concepto de democracia más participativa (Gutmann, 1987), racional y justa y en su apoyo en metodologías mixtas convencionales. El evaluador actúa como un abogado, un negociador, defensor de los principios democráticos, que dialoga y delibera con los diversos participantes, que le presentan y defienden sus intereses. Se plantea como preferente el análisis de las situaciones injustas y la búsqueda de una sociedad verdaderamente democrática.

Un paso más es el que da la *evaluación comunicativa*, que se sustenta en la democracia deliberativa y emancipatoria (Habermas, 1996). Se apoya en las metodologías críticas y se ubica en esferas públicas múltiples. En la práctica, los participantes son los propios gestores de la evaluación, siendo el evaluador un facilitador de la misma. El principal objetivo es el cambio social, actuando y anticipándose a problemas concretos en un contexto determinado.

Como ya hemos señalado con anterioridad, una importante objeción teórico-práctica a este tipo de estrategias es la presión del conservadurismo de los miembros de la institución o programa evaluados cuando se trate de cuestionar principios fundamentales muy asentados en su propia cultura o estructura organizativa. Sin duda, este es un problema que puede frenar ciertas innovaciones, que deben ser alentadas y promovidas con otros esquemas de evaluación y planificación, pero, en un gran porcentaje de las investigaciones evaluativas, las estrategias participativas son extraordinariamente importantes para el éxito de los mismos y para la aplicación de sus sugerencias de cambio.

### 9.3. La toma de decisiones

Hemos podido comprobar que resulta prácticamente imposible escribir una página sobre investigación evaluativa sin referirse al proceso de toma de decisiones o a las decisiones a las que sirve la evaluación. Sin embargo, este tema nos produce cierta sensación de confusión interna al tratar de ubicarlo y analizarlo. A veces, esta especie de confusión se convierte en cierta sensación de culpabilidad profesional, cuando leemos algo como lo que dice Mateo (1990: 95), esto es, «la mayoría de los textos al uso, ventilan el tema señalando su importancia y poco más».

Lo cierto es que la toma de decisiones no es un terreno especialmente explorado por los investigadores evaluativos, aunque podamos encontrar algunos ejemplos interesantes. La mayoría de los trabajos sobre el tema parten de una perspectiva relacionada con la toma de decisiones con base estadística, algo que no encontramos como objeto de tratamiento directo dentro de la investigación evaluativa, aunque sea muy importante para el proceso al que sirve este tipo de investigación.

Una duda importante, y quizás esta sea una de las razones de la relativa escasa atención, es si la toma de decisiones es o no algo en lo que tenga que entrar un investigador evaluativo. Es obvio que se investiga para ese proceso, pero no está tan claro que el investigador deba participar en la toma de decisiones. Nuestra posición personal se inclina hacia un evaluador que ofrece conclusiones valorativas e incluso analiza críticamente alternativas de acción, pero que no debe asumir responsabilidades en la toma de decisiones.

Probablemente, la posición más coherente y acertada sea la de distinguir entre *análisis de decisiones* y *toma de decisiones*, siendo el primero responsabilidad del investigador evaluativo, pero no la toma de decisiones, que debe re-

caer en las distintas audiencias implicadas, en función de su nivel de responsabilidad. Esta parece ser la posición de muchos de los evaluadores que han estudiado el tema (Pitz y McKillip, 1984; Ewell, 1989; Mateo, 1990; De Miguel, 1995). Con esta perspectiva, el análisis de decisiones es un proceso técnico, instrumental, al servicio de una acción, la toma de decisiones, no solo técnica, sino también política y administrativa; el análisis de decisiones proporciona una estructura formal para facilitar el acierto en esta última acción.

La toma de decisiones debe seguir una secuencia que, de acuerdo con Kepner y Tregoe (1989), es la siguiente: *a)* definir objetivos de la decisión, *b)* clasificación conforme a su importancia, *c)* desarrollo de acciones alternativas, *d)* valoración y selección de las alternativas en relación con los objetivos para establecer la decisión tentativa, *e)* análisis crítico de esta y de sus posibles consecuencias y *f)* decisión final y control de sus efectos.

El análisis de decisiones para ayudar en el proceso anterior suele conllevar un trabajo largo y tedioso, y por ello no se utiliza siempre, salvo cuando el problema a resolver es muy complicado, no existe mucha autoridad entre los responsables de las decisiones o se necesita apoyo de justificación o por razones de prestigio.

En este acercamiento al análisis de decisiones por parte de los evaluadores, nos encontramos con dos grandes enfoques o aproximaciones (Scheerens, 1990; De Miguel, 1995). El enfoque *racional* supone la elección entre diversas alternativas claramente definidas en cuanto a sus efectos en función de las metas previstas, mientras que en el enfoque *incremental* se supone que la mayoría de las veces las alternativas no son consecuencia lógica entre medios y fines y se debe contextualizar el proceso a las formas y organizaciones que deben tomar las decisiones.

En la práctica, como ya hemos dicho anteriormente, los evaluadores se han ocupado poco del análisis de este proceso, que, en consecuencia, no tiene quizás el debido desarrollo, y, además, no han solido tener suficiente influencia en dicho proceso. La toma de decisiones, esto es, la utilización de los resultados de las evaluaciones, debe potenciarse considerablemente, pues constituye la verdadera justificación de la importancia de la investigación evaluativa. Como ya hemos señalado, sin utilización ulterior no tiene verdadera justificación la evaluación.

En el ámbito de la educación superior, la toma de decisiones se ha convertido en las últimas décadas en un asunto de creciente interés (Arbizu, 1995), sobre todo por su imbricación con los modelos de gobierno y gestión de las instituciones (De Boer y Goedegebuure, 1995). El auge de la autonomía, la competencia y la rendición de cuentas en los centros de educación superior es causa directa de tal preocupación por los procesos y por la responsabilidad en la toma de decisiones. En este contexto, el análisis comparado nos ofrece un panorama variado en cuanto a jerarquización y centralización de los procesos, según el tipo de instituciones y el tipo de funciones sobre las que hay que decidir. En todo caso, estamos ante un campo de trabajo que necesariamente va a demandar más atención de parte de los investigadores evaluativos.

#### **9.4. Ética y utilización de la evaluación**

El evaluador, que trata habitualmente con objetos de estudio, programas, instituciones, personas, etc., sobre los que pueden incidir determinadas decisiones políticas, a resultas de las conclusiones de sus análisis, actúa en un

terreno intermedio, y muy comprometido, entre los responsables de estas decisiones y los afectados por ellas. En este contexto, es fundamental el respeto por los derechos de unos y de otros, algo que el evaluador debe mantener como garante, a menudo, en situaciones de presión de uno y otro signo.

Una de las primeras obligaciones en la investigación evaluativa es el trato respetuoso y justo de los derechos de los individuos a los que se les aplican los programas, o están implicados en el proceso u objeto de la evaluación, y que proporcionan información sobre los mismos. En tal sentido, Gredler (1996) nos habla del seguimiento de tres principios básicos, que denomina de *beneficio, respeto y justicia*.

El *principio de beneficio* implica dos responsabilidades, maximizar los efectos positivos y evitar el daño (Sieber, 1980). Los estudios deben estar siempre orientados a mejorar el objeto evaluado y la situación de todos los potencialmente afectados, sin discriminaciones de grupos concretos. En cuanto al segundo aspecto, se deben evitar los procedimientos peligrosos, por sus efectos físicos o psicológicos, se debe asegurar el anonimato individual y se debe proteger la confidencialidad. Estos dos últimos aspectos van con frecuencia íntimamente ligados y suelen plantear al evaluador algunos conflictos de intereses no siempre fáciles de resolver. La audiencia y el respeto a la opinión de los implicados suele ser un buen camino para minimizar conflictos.

El *principio de respeto* se refiere a la protección de la autonomía y libertad de las personas en general y el bienestar de las que no gozan de tal autonomía, por edad o por otra situación concreta. El *principio de justicia* nos conduce a tener que pensar en la equidad y la justicia social (House, 1991). La investigación evaluativa,

como práctica social, entraña la ética de la responsabilidad pública y debe ser tremendamente sensible a las necesidades sociales.

Pero el evaluador no solo está obligado a respetar los derechos de los individuos implicados o afectados por el estudio, también tiene obligaciones contraídas con los promotores del mismo y con los responsables de la toma de decisiones. Por ello, a menudo surgen conflictos de obligaciones e intereses que suelen partir del propio promotor del estudio, del mismo evaluador o de la naturaleza conflictiva del objeto evaluado o de las decisiones consiguientes. Stufflebeam (1999) analiza el problema que surge cuando la evaluación produce hallazgos que, por su trascendencia o gravedad, debieran ser conocidos por algún cliente del programa, a pesar de no estar permitido por el contrato de la evaluación. Según su interpretación, los Estándares de la Evaluación (Joint Committee, 1994) permiten al evaluador determinadas acciones en la línea de cumplir con su obligación de publicitar dichos hallazgos entre las audiencias afectadas. Este Comité defiende la tendencia a la neutralidad política en la evaluación a través del estándar que denomina de *viabilidad política*, en el que se establece que «la evaluación debe planificarse y llevarse a cabo anticipando las posiciones de los diversos grupos de interés, para obtener su colaboración y detectar y contrarrestar posibles intentos de mutilación de la evaluación o de sesgo y mal uso de sus resultados» (71).

Stake (1998) incide en la obligación de atender a los múltiples intereses y códigos éticos, también a los de los contrarios y, especialmente, a los de los desfavorecidos, en consonancia con los planteamientos de la corriente postmodernista de la evaluación. Por ello, es crítico con la mayoría de las prácticas evaluadoras recientes, que, en su opinión, atienden básicamente los intereses de los

contratantes de las evaluaciones, descuidando otros intereses y la propia calidad de los programas.

Los estudios que son promovidos para servir a intereses concretos, y no a los intereses generales, no son evaluaciones legítimas; son, simplemente, *pseudoevaluaciones* (Stufflebeam, 1980). Dentro de estas distinguimos entre las controladas políticamente y las inspiradas en las relaciones públicas. En las primeras, el promotor, un responsable político sobre el objeto evaluado, selecciona y controla la información que se debe diseminar, en función de sus intereses políticos, mientras que en las segundas se centra el estudio solo en determinados aspectos o variables que mejoren la imagen social del objeto evaluado. En el primer caso la pseudoevaluación es un instrumento de justificación de una política concreta, y en el segundo de propaganda.

Muy recientemente, House (2008) lanzó una llamada de atención contra el peligro de infectar y poner en serio peligro el campo de la investigación evaluativa con malas prácticas procedentes de un terreno con una buena imagen en cuanto al diseño de sus evaluaciones. Se refería en concreto a los procesos de evaluación de fármacos, que, según este autor, con cierta frecuencia están sesgados al enfatizar excesivamente los efectos positivos del fármaco evaluado y esconder los negativos, incluso manipulando los diseños de investigación para obtener los resultados deseados. No se puede dejar solo en manos de las empresas beneficiadas ni del mercado, señala este autor, un asunto tan decisivo e importante desde el punto de vista social. En este y en todos los asuntos sociales, los responsables públicos y los propios evaluadores deben luchar frontalmente contra las pseudoevaluaciones y contra las malas evaluaciones y las evaluaciones manipuladas (*misevaluations*).

El mismo evaluador puede ser una fuente de conflictos cuando sus valores personales son muy contrarios, o son muy favorables, al objeto mismo de la evaluación. Así, por ejemplo, alguien implicado personalmente en la defensa o en la crítica de un programa educativo concreto es muy difícil que pueda evaluarlo sin generar dudas sobre su trabajo.

Finalmente, el propio objeto de la evaluación puede ser conflictivo y le puede plantear problemas de naturaleza ética al investigador. Un caso habitual es cuando existen defensores y detractores del programa evaluado. Entonces, el evaluador debe dar audiencia a todas las partes en conflicto y articular un sistema armonioso de confluencia de criterios y de elaboración de conclusiones y sugerencias.

Un segundo tipo de situación conflictiva se produce cuando se evalúan intervenciones sociales que no tienen el soporte de la investigación. En tal caso, el evaluador debe ser extremadamente riguroso en el análisis de resultados de la misma, contrastando los efectos en diferentes contextos y comparando tal realidad con la que se da en situaciones sin intervención. En suma, en esta circunstancia el evaluador debe dejar constancia de su condición de investigador experimental.

Christie y Alkin (1999) señalan que la mayoría de la literatura sobre ética y evaluación se centra en el evaluador y, más en concreto, en sus principios y en su práctica. En otras palabras, se resalta el peligro de las malas evaluaciones (*misevaluations*), mal orientadas, sesgadas, incompletas o débiles metodológicamente. Pero ellos quieren enfatizar los peligros del mal uso y abuso (*misuse*) de las evaluaciones que hacen otras personas, responsables políticos, gestores, etc., que son más difíciles de controlar y combatir. Estos autores ofrecen una síntesis clasificato-

ria de los usos y malos usos que suelen darse con los programas de evaluación.

Cruzando las dimensiones de (*uso-no uso*) y (*uso correcto-mal uso*), Cousins (2004) establece cuatro cuadrantes o categorías con relación a la utilización de las evaluaciones. Estas categorías son las siguientes:

- 1) *Utilización ideal*, en la que se recogen los usos instrumentales, conceptuales y de persuasión o apoyo antes comentados.
- 2) *Mala utilización (misuse)*, bien por incompetencia o incomprensión de los resultados de la evaluación, bien por manipulación intencionada de los mismos.
- 3) *No utilización injustificada*, desconsiderando de manera abusiva o eliminando los resultados de la evaluación.
- 4) *No utilización justificada*, cuando aparecen razones de diverso tipo que aconsejan no tener en cuenta los resultados de la evaluación.

Aunque no resulte fácil, por su propia situación en el proceso de toma de decisiones, el evaluador no solo debe intentar evitar que sus resultados sean utilizados de manera injustificada (categoría 3), por desgracia una situación muy frecuente, sino que debe hacer todo lo posible por evitar el mal uso de su trabajo (categoría 2), sin duda el problema más peligroso.

House (1995) nos señala cinco *falacias éticas* que resaltan cinco problemas de los que el evaluador está obligado a huir. Estas falacias son las siguientes:

- *Clientelismo*. La falacia de que hacer cualquier cosa que el cliente demande o que le beneficie es éticamente correcto.
- *Contractualismo*. La falacia de que el evaluador está obligado a seguir el contrato escrito de manera rígida, incluso si ello va en detrimento del bien público.

- *Metodologismo*. La creencia de que el seguimiento de métodos de investigación aceptables asegura la ética en la conducta del investigador, aunque en tales metodologías puedan estar implícitos ciertos dilemas éticos.
- *Relativismo*. La falacia de que todas las opiniones recogidas por el evaluador entre diversos grupos participantes son igualmente importantes, como si no hubiera razones para establecer prioridades según el nivel de implicación y relevancia ante el objeto de evaluación de los grupos informantes.
- *Pluralismo/elitismo*. La falacia de dar más prioridad a voces poderosas, no porque lo merezcan, sino porque tienen más prestigio o potencia que otras voces implicadas.

Como es lógico, la investigación evaluativa que es verdaderamente útil y rigurosa se hace al margen de estas falacias que ponen en peligro la ética en las evaluaciones y en sus usos, algo que, como hemos analizado, es fundamental para el desarrollo y credibilidad de un tipo de investigación aplicada cada vez más relevante para asegurar y acrecentar la calidad de nuestras políticas sociales en general y, en definitiva, de nuestra sociedad. Una investigación que, como señala Chelimsky (2008), debe buscar en todo momento el ajuste entre la necesaria independencia científica y las justas demandas de una sociedad democrática.

10  
REFLEXIÓN PARA FINALIZAR

En el presente trabajo hemos recorrido el camino y los procesos seguidos por la investigación evaluativa hasta consolidarse como una disciplina transversal de gran pujanza, especialmente en el campo de las ciencias sociales, así como una orientación investigadora de creciente utilización en múltiples ámbitos del conocimiento y del desarrollo social.

Partiendo de la definición actual de investigación evaluativa, hemos enfatizado la función de esta al servicio del cambio social y, en concreto, al servicio de la mejora social. Hemos visto que la evaluación resulta ser un elemento imprescindible de planificación, de gestión y de mejora de la sociedad, pero también que esta función de cambio y mejora no siempre ha sido resaltada al hablar de evaluación como lo es en la actualidad, pues es a partir de los años sesenta cuando empieza a consolidarse esta función como verdaderamente relevante, rompiendo y extendiendo el restrictivo marco del control, de la rendición de cuentas, de la fiscalización y de la simple valoración con fines comparativos. El gran objetivo es mejorar la sociedad y, más en concreto, la calidad de vida de los ciudadanos, pero este es un concepto complejo, dinámico, multidimensional, necesita-

do de contextualización, que requiere análisis y tratamientos rigurosos y reposados.

El análisis de la utilidad y el uso de la evaluación constituye un bloque central y extenso en este trabajo por varias razones. En primer lugar, porque se trata de uno de los mayores problemas con el que hoy se encuentran los evaluadores como profesionales; en segundo lugar, porque no siempre se le presta la debida atención en la literatura especializada; y, en tercer lugar, porque el futuro de la investigación evaluativa como disciplina académica y como profesión relevante está directamente relacionado con y condicionado por el uso y la utilización que se haga de las evaluaciones y de sus resultados. Si las investigaciones evaluativas no son útiles para el cambio y la mejora social y si no se utilizan bien, nuestra tesis es que son un esfuerzo poco relevante y, por lo tanto, no deberían plantearse. Pero sabemos y defendemos que, si se utilizan bien, son una herramienta de especial relevancia estratégica.

En el trabajo presentamos distintos modelos de utilización, usos de los resultados, así como distintas fórmulas y estrategias facilitadoras de esta utilización. Entramos también en el debate que ha surgido recientemente alrededor de los conceptos de uso y de influencia. También tratamos el problema de la toma de decisiones, al que siempre se hace referencia, pero que raramente se trata. Defendemos la implicación del evaluador en el análisis de las decisiones alternativas, pero pensamos que su papel no debe ser especialmente relevante en el proceso de toma de decisiones propiamente dicho. Concluimos este bloque con una defensa nítida de las estrategias participativas en la evaluación como estímulo y garantía de uso y buena utilización de sus resultados, aunque reconocemos alguna de sus potenciales debilida-

des. Presentamos las distintas fórmulas de participación que se plantean o pueden plantearse.

El trabajo concluye con una llamada de atención sobre la utilización ética de las evaluaciones, con la presentación de los principios que deben seguirse, con el análisis de algunos usos y abusos que aparecen con cierta frecuencia, y, finalmente, presentamos algunas falacias sobre estrategias práctico-metodológicas de las que el buen evaluador debe huir.

Como idea final, queremos hacer una llamada a nuestros departamentos universitarios para que se impliquen cada vez más en investigaciones evaluativas bien fundamentadas y desarrolladas, que mejoren la realidad social y las políticas en su ámbito de conocimiento. Ahora bien, es importante señalar que la investigación evaluativa no puede ni debe sustituir a la investigación fundamental, básica y aplicada; simplemente, debe convivir con ellas dentro del rango de tareas de nuestros científicos, manteniendo y respondiendo cada orientación investigadora a sus retos específicos para el desarrollo del conocimiento y de la sociedad.



## REFERENCIAS

- ALKIN, M., 1969, Evaluation theory development, *Evaluation Comment*, 2, 1, 2-7.
- ALKIN, M. C., 1991, *Debates on Evaluation*, Sage, Newbury Park, Ca.
- ALVIRA, F., 1983, Perspectiva cualitativa-Perspectiva cuantitativa en la metodología sociológica, *Revista Española de Investigaciones Sociológicas*, 22, 53-75.
- AMERICAN EVALUATION ASSOCIATION, 1995, Guiding principles for evaluators, en Shadish, W. R. y otros (eds.), *New directions for program evaluation*, Jossey-Bass, San Francisco, 19-26. Ver también: <<http://www.eval.org/>>.
- AMERICAN PSYCHOLOGICAL ASSOCIATION, 1954, *Technical recommendation for psychological tests and diagnostic techniques*, APA, Washington, DC.
- ANDERSON, G., 1990, *Fundamentals of educational research*, The Falmer Press, London.
- ANDERSON, S. B. y BALL, S., 1983, *The profession and practice of program evaluation*, Jossey-Bass, San Francisco, Ca.
- ARBIZU, F. (ed.), 1995, *La toma de decisiones. Un elemento clave en la mejora universitaria*, ICE, Universidad del País Vasco.
- ARNAL, J. y otros, 1992, *Investigación educativa. Fundamentos y metodología*, Labor, Barcelona.
- BERK, R. A. (ed.), 1981, *Educational evaluation methodology: The state of the art*, Johns Hopkins University Press, Baltimore.

- BERK, R. A. y ROSSI, P. H., 1990, *Thinking about program evaluation*, Sage, Newbury Park, Ca.
- BOGDAN, R. y BIKLEN, S., 1982, *Qualitative research for education*, Allyn and Bacon, London.
- BREDO, E., 2006, Philosophies of Educational Research, en Green, J. L., Camilli, G. y Elmore, P. B., *Handbook of Complementary Methods in Education Research*, AERA, Lawrence Erlbaum, London, 3-31.
- CAJIDE, J., 1992, La investigación cualitativa: tradiciones y perspectivas contemporáneas, *Bordón*, 44, 4, 357-373.
- CALDERON, A. J., 2004, Institutional Research at RMIT. A Case study. Ponencia presentada en el *26th EAIR Forum*, Barcelona, 5-8 de septiembre.
- CAMPBELL, D. T. y STANLEY, J. C., 1963, *Experimental and quasi-experimental designs for research*, Rand McNally, Chicago.
- CARR, W. y KEMMIS, S., 1983, *Becoming critical: Knowing through action research*, Deakin University, Victoria, Australia.
- CASTILLO, S. y GENTO, S., 1995, Modelos de evaluación de programas educativos, en Medina, A. y Villar, L. M. (coords.), *Evaluación de programas educativos, centros y profesores*, Universitat, Madrid, 25-69.
- CHELIMSKY, E., 1998, The role of experience in formulating theories of evaluation practice, *American Journal of Evaluation*, 19, 1, 35-55.
- CHELIMSKY, E., 2008, A Clash of Cultures: Improving the «Fit» Between Evaluative Independence and the Political Requirements of a Democratic Society, *American Journal of Evaluation*, 29, 4, 400-415.
- CHRISTIE, C. A., 2003, The practice-theory relationship in evaluation, *New Directions for Program Evaluation*, 97, Jossey-Bass, San Francisco, Ca.
- CHRISTIE, C. A. y ALKIN, M. C., 1999, Further reflections on evaluation misutilization, *Studies in Educational Evaluation*, 25, 1, 1-10.
- CHRISTIE, C. A., ROSS, R. M. y KLEIN, B. M., 2004, Moving toward collaboration by creating a participatory internal-external evaluation team: a case study, *Studies in Educational Evaluation*, 30, 2, 125-134.

- CLARKE, A. y DAWSON, R., 1999, *Evaluation Research*, Sage, London.
- COLÁS, P. y REBOLLO, M. A., 1993, *Evaluación de programas. Una guía práctica*, Kronos, Sevilla.
- CONVERSE, J. M., 1986, *Survey research in the US: Roots and emergence 1890-1960*, University of California Press, Los Angeles.
- COOK, T. D. y REICHHARDT, C. S., 1986, *Métodos cualitativos y cuantitativos en investigación educativa*, Morata, Madrid.
- COUSINS, J. B., 2004, Commentary: Minimizing evaluation misuse as principled practice, *American Journal of Evaluation*, 25, 3, 391-397.
- CRONBACH, L. J., 1957, The two disciplines of scientific psychology, *American Psychologist*, 12, 671-684.
- CRONBACH, L. J., 1963, Course improvement through evaluation, *Teachers College Record*, 64, 672-683.
- CRONBACH, L. J., 1975, Beyond the two disciplines of scientific psychology, *American Psychologist*, 30, 116-127.
- CRONBACH, L. J., 1982, *Designing evaluations of educational and social programs*, Jossey-Bass, San Francisco.
- CRONBACH, L. J. y otros, 1980, *Toward reform of program evaluation*, Jossey-Bass, San Francisco.
- CRONBACH, L. J. y SUPPES, P., 1969, *Research for tomorrow's schools: Disciplined inquiry for education*, MacMillan, New York.
- DE BOER, H. y GOEDEGEBUURE, L., 1995, Decision-making in higher education. A comparative perspective, en *XVII EAIR Forum*, Zurich, 20 págs.
- DE LA ORDEN, A., 1985, Investigación evaluativa, en De la Orden, A. (ed.), *Investigación educativa. Diccionario de Ciencias de la Educación*, Anaya, Madrid, 133-137.
- DE MIGUEL, M., 1995, La evaluación de las instituciones universitarias y la toma de decisiones, en Arbizu, F. (ed.), *La toma de decisiones*, Unidad de Evaluación, ICE-Universidad del País Vasco, 77-142.
- DE MIGUEL, M., ESCUDERO, T. y RODRÍGUEZ, S., 1998, Spanish universities' quality: the incentive of external evaluation, *Quality in Higher Education*, 4, 2, 199-206.
- DONALDSON, S. I., 2007, *Program theory-driven evaluation science*, Erlbaum, Mahwah, NJ.

- DONALDSON, S. I. y LIPSEY, M. W., 2006, Roles for theory in contemporary evaluation practice: Developing practical knowledge, en Shaw, I., Greene, J. C. y Mark, M. M. (eds.), *The Handbook of Evaluation: Policies, programs, and Practices*, Sage, London, 56-75.
- DONALDSON, S. I. y SCRIVEN, M., 2003, Diverse visions for evaluation in the new millenium, en Donaldson, S. I. y Scriven, M. (eds.), *Evaluating social programs and problems: Vision for the new millenium*, Erlbaum, Mahwah, NJ, 3-16.
- EDWARDS, W. y otros, 1975, A decision-theoretic approach to evaluation research, en Struening, E. L. y Guttentag, M. (eds.), *Handbook of Evaluation Research*, Beverly Hills, Ca, 139-181.
- EISNER, E. W., 1975, *The perceptive eye: Toward the reformation of educational evaluation*, Stanford Evaluation Consortium, Stanford, Ca.
- EISNER, E. W., 1985, *The art of educational evaluation*, The Falmer Press, London.
- EISNER, E. W., 1988, The primacy of experience and the politics of methods, *Educational Researcher*, 17, 5, 15-20.
- ESCUADERO, T., 1980, ¿Se pueden evaluar los centros educativos y sus profesores?, *Educación Abierta*, 10, ICE-Universidad de Zaragoza.
- ESCUADERO, T., 1993, Enfoques modélicos en la evaluación de la enseñanza universitaria, en *Actas de las III Jornadas Nacionales de Didáctica Universitaria «Evaluación y Desarrollo Profesional»*, Servicio de Publicaciones, Universidad de Las Palmas, 5-59.
- ESCUADERO, T., 1996, *Proyecto docente e investigador*, Universidad de Zaragoza.
- ESCUADERO, T., 2000, Evaluación de centros e instituciones educativas: las perspectivas del evaluador, en González, D. y otros (coords.), *Innovación en la escuela y mejora de la calidad educativa*, Grupo Editorial Universitario, Granada, 57-76.
- ESCUADERO, T., 2002, Evaluación institucional: algunos fundamentos y razones, en Álvarez, V. y Lázaro, A., *Calidad de las Universidades y Orientación Universitaria*, Aljibe, Málaga, 103-138.

- ESCUADERO, T., 2003, Desde los tests hasta la investigación evaluativa actual. Un siglo, el xx, de intenso desarrollo de la evaluación en educación. [English version: From tests to current evaluative research. One century, the xxth, of intense development of evaluation in education]. *RELIEVE*, vol. 9, n.º 1., <<http://www.uv.es/RELIEVE/>>.
- ESCUADERO, T., 2005-2006, Claves identificativas de la investigación evaluativa: análisis desde la práctica, *Contextos Educativos. Revista de Educación*, 8-9, 179-199.
- ESCUADERO, T., 2006, Evaluación y mejora de la calidad en educación, en Escudero, T. y Correa, A. D., *Investigación en innovación educativa: algunos ámbitos relevantes*, La Muralla, Madrid, 269-325.
- ESCUADERO, T., 2007, Evaluación institucional de la calidad universitaria en España: Breve pero interesante historia, *Anuario de Pedagogía*, 9, 103-115.
- ESCUADERO, T., 2009, Some relevant topics in educational evaluation research, en Asorey, M., García Esteve, J. V., F. Rañada, M. y Sesma, J., *Mathematical Physics and Field Theory. Julio Abad, in memoriam*, Prensas Universitarias de Zaragoza, Zaragoza, 223-230.
- ESCUADERO, T., PINO, J. L. y RODRÍGUEZ, C., 2010, Evaluación del profesorado universitario para incentivos individuales: revisión metaevaluativa, *Revista de Educación*, 351, 513-537.
- EVALUATION THESAURUS, 1991, Sage, Newbury Park, Ca.
- EWELL, P. T., 1989, *Enhancing information use in decision making*, New Directions for Institutional Research, Jossey-Bass, San Francisco.
- FETTERMAN, D. M., 1984, *Ethnography. Step by step*, Sage, London.
- FETTERMAN, D. M., 2000, *Foundations of empowerment evaluation*, Sage, Thousand Oaks, Ca.
- FISHMAN, D. B., 1988, Pragmatic behaviorism: Saving and nurturing the baby, en Fishman, D. B. y otros (eds.), *Paradigms in behavior therapy: Present and promise*, Springer, New York, 254-293.
- FISHMAN, D. B., 1991, An introduction to the experimental versus the pragmatic paradigm in evaluation, *Evaluation and Program Planning*, 14, 4, 353-363.

- GAGE, N. L., 1989, The paradigm wars and their aftermath. A «historical» sketch of research on teaching since 1989, *Educational Researcher*, 18, 7, 4-10.
- GEPHART, W. J. y POTTER, J., 1976, The generalizability problem in evaluation and a borrowed solution, *CEDR Quarterly*, Summer, 9-11.
- GLASER, R., 1963, Instructional technology and the measurement of learning outcomes: some questions, *American Psychologists*, 18, 519-521.
- GLASS, G. V. y otros, 1975, *Design and analysis of time-series experiments*, Colorado Associated University Press, Boulder, Colorado.
- GLASS, G. V. y otros, 1981, *Meta-analysis in social research*, Sage, Beverly Hills, Ca.
- GLASS, G. V. y WORTHEN, B. R., 1971, Evaluation and research: similarities and differences, *Curriculum Theory Network*, Fall, 149-165.
- GRASSO, P. G., 1999, Meta-evaluation of an evaluation of reader focused writing for the Veterans Benefits Administration, *American Journal of Evaluation*, 20, 3, 355-371.
- GRASSO, P. G., 2003, What makes an evaluation useful? Reflections from experience in large organizations, *American Journal of Evaluation*, 24, 4, 507-514.
- GREDLER, M. E., 1996, *Program evaluation*, Merrill/Prentice-Hall, Englewood Cliffs, NJ.
- GUBA, E. G. (ed.), 1990, *The paradigm dialogue*, Sage, Newbury Park, Ca.
- GUBA, E. G. y LINCOLN, Y. S., 1982, Epistemological and methodological bases of naturalistic inquiry, *Educational Communication and Technology Journal*, 30, 233-252.
- GUBA, E. G. y LINCOLN, Y. S., 1989, *Fourth Generation Evaluation*, Sage, Newbury Park, Ca.
- GUTMANN, A., 1987, *Democratic evaluation*, University Press, Princeton, NJ.
- HABERMAS, J., 1971, *Knowledge and human interest*, Beacon Press, Boston.
- HABERMAS, J., 1973, *Theory and practice*, Beacon Press, Boston.

- HABERMAS, J., 1996, Three normative models of democracy, en Benhabib, S. (ed.), *Democracy and difference*, Princeton University Press, Princeton, NJ, 21-30.
- HAMBLETON, R. K., 1985, Criterion-referenced measurement, en *Encyclopedia of Educational Research*, McMillan, New York.
- HENRY, G. T., 2003, Influential evaluations, *American Journal of Evaluation*, 24, 4, 515-524.
- HENRY, G. T. y MARK, M. M., 2003, Beyond use: Understanding evaluation s influence on attitudes and actions, *American Journal of Evaluation*, 24, 3, 293-314.
- HERZOG, E., 1959, *Some Guidelines for Evaluative Research*, U.S. Department of Health, Education, and Welfare, Washington, DC.
- HOUSE, E. R., 1973, Politics as usual, en House, E. R. (ed.), *School Evaluation. The Politics and the Process*, McCutchan, Berkeley, 329-331.
- HOUSE, E. R., 1989, *Evaluating with validity*, Sage, Newbury Park, Ca.
- HOUSE, E. R., 1991, Evaluation and social justice: Where are we?, en McLaughlin, M. y Phillips, D. C. (eds.), *Evaluation and education: At quarter century. Nineteenth yearbook of the National Society for the Study of Education, Part II*, University of Chicago Press, Chicago, 233-247.
- HOUSE, E. R., 1993, *Profesional Evaluation: Social Impact and Political Consequences*, Sage, Newbury Park, Ca.
- HOUSE, E. R., 1995, Principled evaluation: A critique of the AEA Guiding Principles, en Shadish, W. R., Newman, D. L., Scheirer, M. A. y Wye, C. (eds.), *Guiding principles for evaluators*, New Directions for Program Evaluation, n.º 66, Jossey-Bass, San Francisco, 27-34.
- HOUSE, E. R., 2008, Blowback. Consequences of Evaluation for Evaluation, *American Journal of Evaluation*, 29, 4, 416-426.
- HOWE, K. R., 1985, Two dogmas of educational research, *Educational Researcher*, 14, 8, 10-18.
- HOWE, K. R., 1988, Against the qualitative-quantitative incompatibility thesis or dogmas die hard, *Educational Researcher*, 17, 9, 10-16.

- HUSEN, T., 1988, Research paradigms in education, en Keeves, J. P. (ed.), *Educational Research, Methodology and Measurement: An International Handbook*, Pergamon Press, Oxford.
- ISAAC, S. y MICHAEL, W. B., 1984, *Handbook in research and evaluation*, Edit, San Diego, Ca.
- JACOB, E., 1987, Qualitative research traditions: A review, *Review of Educational Research*, 57, 1-50.
- JANG, S., 2000, The appropriateness of Joint Committee standards in non-western settings: A case study in South Korea, en Russon, C. (ed.), *The program evaluation standards in international settings*, The Evaluation Center Occasional Papers Series n.º 17, Kalamazoo, Mi, 41-59.
- JOINT COMMITTEE on Standards for Educational Evaluation, 1981, *Standards for evaluations of educational programs, projects, and materials*, McGraw-Hill, New York.
- JOINT COMMITTEE on Standards for Educational Evaluation, 1988, *The personnel evaluation standards*, Sage, Newbury Park, Ca.
- JOINT COMMITTEE on Standards for Educational Evaluation, 1994, *The program evaluation standards*, Sage, Thousand Oaks, Ca.
- JOINT COMMITTEE on Standards of Educational Evaluation, 2003, *The student evaluation standards*, Corwin, Thousand Oaks, Ca.
- KEEFE, J., 1994, School evaluation using the CASE-IMS model and improvement process, *Studies in Educational Evaluation*, 20, 1, 55-67.
- KEEVES, J. P. (ed.), 1988, *Educational Research Methodology, and Measurement: An International Handbook*, Pergamon Press, Oxford.
- KEPNER, H. y TREGOE, B., 1989, *El directivo racional*, Facultad de Ciencias Económicas y Empresariales, Universidad Autónoma de Barcelona.
- KIRKHART, K., 2000, Reconceptualizing evaluation use: An integrated theory of influence, en Caracelli, V. y Preskill, H. (eds.), *The expanding scope of evaluation use. New Directions for Evaluation*, n.º 88, Jossey-Bass, San Francisco, Ca, 5-24.

- KISH, L., 1987, *Statistical Design for Research*, John Wiley & Sons, New York.
- KUHN, T. S., 1962, *The structure of scientific revolutions*, University of Chicago Press, Chicago.
- KUPFERSMID, J., 1988, Improving what is published. A model in search of an editor, *American Psychologist*, 43, 635-642.
- LECOMPTE, M. D., 1994, Sensible matchmaking: Qualitative research design and the program evaluation standards, *Journal of Experimental Education*, 63, 1, 29-43.
- LEVITON, L. C., 2003, Evaluation use: Advances, challenges and applications, *American Journal of Evaluation*, 24, 4, 525-535.
- LOCKE, J., 1988, *Two treatises of government*, Cambridge University Press, New York.
- MACDONALD, B., 1976, Evaluation and the control of education, en Tawney, D. (ed.), *Curriculum evaluation today: trends and implications*, McMillan, London, 125-136.
- MACDONALD, B., 1987, Evaluation and the control of education, en Murphy, R. y Torrance, H. (eds.), *Evaluating education: issues and methods*, Harper and Row, London, 36-48.
- MADAUS, G. F. y otros, 1991, *Evaluation Models. Viewpoints on Educational and Human Services Evaluation*, Kluwer-Nijhoff, Hingham, Mass.
- MAGUIRE, T. O., 1990, On methodologic sects in educational research, *The Alberta Journal of Educational Research*, XXXVI, 4, 297-298.
- MARK, M. M., 2003, Toward a integrative view of the theory and practice of program and policy evaluation, en Donaldson, S. I. y Scriven, M. (eds.), *Evaluating social programs and problems: Vision for the new millenium*, Erlbaum, Mahwah, NJ, 183-204.
- MATEO, J., 1990, La toma de decisiones en el contexto de la evaluación educativa, *Revista de Investigación Educativa*, 8, 16, 95-112.
- MATEO, J. y otros, 1993, *La evaluación en el aula universitaria*, Material de cursos, ICE-Universidad de Zaragoza.
- MAY, H., 2004, Making statistics more meaningful for policy research and program evaluation, *American Journal of Evaluation*, 25, 4, 525-540.

- McCLINTOCK, C., 2003, Commentary: The evaluator as scholar/practitioner/change agent, *American Journal of Evaluation*, 24, 1, 91-96.
- McMILLAN, J. H. y SCHUMACHER, S., 1984, *Research in education*, Little, Brown, Canada.
- MERTENS, D. M., 1999, Inclusive evaluation: Implications of transformative theory for evaluation, *American Journal of Evaluation*, 20, 1, 1-14.
- MORIN, A., 1985, Critères de scientificité de la recherche-action, *Revue des Sciences de l'Éducation*, 11, 1, 31-49.
- NEVO, D., 1989, The conceptualization of educational evaluation: An analytical review of the literature, en House, E. R. (ed.), *New directions in educational evaluation*, The Falmer Press, London, 15-29.
- NISBET, J., 1980, Educational research—the state of the art, en Dockrell, W.B. y Hamilton, D. (eds.), *Rethinking education research*, Hodder and Stoughton, London.
- NISBET, J., 1988, Policy-oriented research, en Keeves, J. P. (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*, Pergamon Press, Oxford, 139-146.
- NORRIS, N., 1993, *Understanding educational evaluation*, Kogan Page/CARE, School of Education, University of East Anglia, London.
- OBERLE, K., 1991, Paradigm Wars: Who's fighting; who's winning?, *The Alberta Journal of Educational Research*, XXXVII, 1, 87-97.
- OWEN, J. M. y ROGERS, P. J., 1999, *Program evaluation. Forms and approaches*, Sage, London.
- OWENS, T. R., 1973, Educational evaluation by adversary proceedings, en House, E. R. (comp.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.
- PARLETT, M. y HAMILTON, D., 1977, Evaluation as illumination: A new approach to the study of innovative programmes, en Hamilton D. y otros (eds.), *Beyond the numbers game*, MacMillan, London.
- PATTON, M. Q., 1984, Data collection: options, strategies, and cautions, en Rutman, L. (ed.), *Evaluation research methods: A basic guide*, Sage, Beverly Hills, Ca.

- PATTON, M. Q., 1997, *Utilization focused evaluation*, Sage, Thousand Oaks, Ca.
- PERRIN, B., 2001, Commentary: Making yourself—and evaluation—useful, *American Journal of Evaluation*, 22, 2, 252-259.
- PITZ, C. F. y MCKILLIP, J., 1984, *Decision analysis for programs evaluations*, Sage, Beverly Hills, Ca.
- POLKINGHORNE, D., 1983, *Methodology for the Human Sciences: Systems of Inquiry*, University of New Jersey Press, New York.
- POPHAM, W. J., 1980, *Problemas y técnicas de la evaluación educativa*, Anaya, Madrid.
- POPKWITZ, T., 1984, *Paradigm and ideology in educational research*, The Falmer Press, London.
- POSAVAC, E. J. y CAREY, R. G., 1989, *Program Evaluation. Methods and Case Studies*, Prentice Hall, Englewood Cliffs, NJ.
- PROVUS, M., 1971, *Discrepancy evaluation. For educational program improvement and assessment*, McCutchan, Berkeley, Ca.
- REA-DICKINS, P., 1994, Evaluation and english language teaching, *Language Teaching*, 27, 2, 71-91.
- ROGERS, W. T., 1994, The treatment of measurement issues in the revised program evaluation standards, *Journal of Experimental Education*, 63, 1, 13-28.
- ROSSI, P. H. y FREEMAN, H., 1993, *Evaluation: A Systematic Approach*, Sage, Beverly Hills, Ca.
- RUSSELL, N. y WILLINSKY, J., 1997, Fourth generation educational evaluation: The impact of a post-modern paradigm on school based evaluation, *Studies in Educational Evaluation*, 23, 3, 187-199.
- RUTMAN, L. (ed.), 1984, *Evaluation research methods: A base guide*, Sage, Beverly Hills, Ca.
- RUTMAN, L. y MOWBRAY, G., 1983, *Understanding program evaluation*, Sage Beverly Hills, Ca.
- RYAN, K. E., 2004, Serving public interests in educational accountability: Alternative approaches to democratic evaluation, *American Journal of Evaluation*, 25, 4, 443-460.
- SANDERS, J. R., 1994, The process of developing national standards that meet ANSI guidelines, *Journal of Experimental Education*, 63, 1, 5-12.

- SANDERS, J. R., 1995, Standards and principles, en Shadish, W. R. y otros (eds.), *New directions for program evaluation*, Jossey-Bass, San Francisco, 47-52.
- SCHEERENS, J., 1990, Beyond decision-oriented evaluation, en Walberg, H. y Haertel, G., *The International Encyclopedia of Educational Evaluation*, Pergamon Press, Oxford, 35-40.
- SCHEERENS, J., 2004, The evaluation culture, *Studies in Educational Evaluation*, 30, 2, 105-124.
- SCHWANDT, T. A. y HALPERN, E. S., 1988, *Linking Auditing and Metaevaluation. Enhancing Quality in Applied Research*, Sage, Newbury Park, Ca.
- SCHWARTZ, R. y MAYNE, J., 2005, Assuring the quality of evaluative information: theory and practice, *Evaluation and Program Planning*, 28, 1, 1-14.
- SCRIVEN, M., 1967, The methodology of evaluation, en *Perspectives of Curriculum Evaluation* (AERA Monograph 1), Rand McNally, Chicago, 39-83.
- SCRIVEN, M., 1969, An introduction to meta-evaluation, *Educational Product Report*, 2, 36-38.
- SCRIVEN, M., 1973, Goal-free evaluation, en House, E. R. (ed.), *School evaluation: The politics and process*, McCutchan, Berkeley, Ca, 319-328.
- SCRIVEN, M., 1974, Prose and cons about goal-free evaluation, *Evaluation Comment*, 3, 1-4.
- SCRIVEN, M., 1991, *Evaluation Thesaurus*, Sage, Newbury Park, Ca.
- SCRIVEN, M., 1991a, *Duties of the teacher*, Center for Research on Educational Accountability and Teacher Evaluation, Kalamazoo, Mi.
- SCRIVEN, M., 1994, Evaluation as a discipline, *Studies in Educational Evaluation*, 20, 1, 147-166.
- SCRIVEN, M., 1998, Minimalist theory: The least theory that practice require, *American Journal of Evaluation*, 19, 1, 57-70.
- SCRIVEN, M., 2000, The logic and methodology of checklists, <[www.wmich.edu/evalctr/checklists/](http://www.wmich.edu/evalctr/checklists/)>.
- SCRIVEN, M., 2003, Evaluation in the new millenium: The transdisciplinary vision, en Donaldson, S. I. y Scriven, M. (eds.), *Eva-*

- uating *Social Programs and Problems. Visions for the New Millennium*, Lawrence Earlbaum Associates, Mahwah, NJ, 19-42.
- SHADISH, W. R., 1998, Evaluation theory is who we are, *American Journal of Evaluation*, 19, 1, 1-19.
- SHADISH, W. R., COOK, T. D. y LEVITON, L. C., 1991, *Foundations of Program Evaluation: Theories of Practice*, Sage, Newbury Park, Ca.
- SHULMAN, L., 1988, Disciplines of inquiry in education: An overview, en Jaeger, R., *Complementary methods for research in education*, AERA, Washington, 3-17.
- SIEBER, J., 1980, Being ethical: Professional and personal decisions in program evaluation, en Perloff, R. y Perloff, E. (eds.), *Values, ethics, and standards in evaluation. New Directions for Program Evaluation*, 7, 51-61.
- SMITH, E. R. y TYLER, R. W., 1942, *Appraising and recording student progress*, Harper & Row, New York.
- SMITH, J. K., 1983, Quantitative versus qualitative research: An attempt to clarify the issue, *Educational Researcher*, 12, 3, 6-13.
- SMITH, J. K. y HESHUSIUS, L., 1986, Closing down the conversation: the end of the quantitative-qualitative debate among educational inquirers, *Educational Researcher*, 15, 1, 4-12.
- SMITH, N. L. y HAVER, D. M., 1990, The applicability of selected evaluation models to evolving investigative designs, *Studies in Educational Evaluation*, 16, 3, 489-500.
- SOLTIS, J., 1984, On the nature of educational research, *Educational Researcher*, 10, 5-10.
- STACHER, B. M. y DAVIS, W. A., 1990, *How to Focus on Evaluation*, Sage, Newbury Park, Ca.
- STAKE, R. E., 1967, The countenance of educational evaluation, *Teacher College Record*, 68, 523-540.
- STAKE, R. E., 1975, *Evaluating the arts in education: A responsive approach*, Merrill, Columbus, Ohio.
- STAKE, R. E., 1998, A When policy is merely promotion, by what ethic lives an evaluator, *Studies in Educational Evaluation*, 24, 2, 203-212.
- STAKE, R., 2006, *Evaluación comprensiva y evaluación basada en estándares*, Graó, Barcelona.

- STRAW, R. B. y COOK, T. D., 1985, Metaevaluation, en Husen, T. y Posththlethwaize, D. (eds.), *International Encyclopedia of Education*, Pergamon, Oxford.
- STRONGE, J. H. y HELM, V. M., 1991, *Evaluating professional support personnel in educational settings*, Sage, Newbury Park, Ca.
- STUFFLEBEAM, D. L., 1980, Alternative approaches to evaluation, en Madaus, G. y otros (eds.), *Evaluation Models*, Kluwer-Nijhoff, Boston.
- STUFFLEBEAM, D. L., 1994, Introduction: Recommendations for improving evaluations in U. S. public schools, *Studies in Educational Evaluation*, 20, 1, 3-21.
- STUFFLEBEAM, D. L., 1998, Conflicts between standards-based and postmodernist evaluations: Toward rapprochement, *Journal of Personnel Evaluation in Education*, 12, 3, 287-296.
- STUFFLEBEAM, D. L., 1999, Using professional standards to legally and ethically release evaluation findings, *Studies in Educational Evaluation*, 25, 4, 325-334.
- STUFFLEBEAM, D. L., 2000, Guidelines for developing evaluation checklists, <[www.wmich.edu/evalctr/checklists/](http://www.wmich.edu/evalctr/checklists/)>.
- STUFFLEBEAM, D. L., 2001, The metaevaluation imperative, *American Journal of Evaluation*, 22, 2, 183-209.
- STUFFLEBEAM, D. L., 2004, A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards, *American Journal of Evaluation*, 25, 1, 99-102.
- STUFFLEBEAM, D. L. y otros, 1970, *Educational evaluation and decision-making*, Evaluation Center, Ohio State University, Columbus, Ohio.
- STUFFLEBEAM, D. L. y SHINKFIELD, A. J., 1987, *Evaluación sistemática. Guía teórica y práctica*, Paidós/MEC, Barcelona.
- SUCHMAN, E. A., 1967, *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*, Russell Sage Foundation, New York.
- SUSSER, M., 1975, Epidemiological models, en Struening, E. y Guttentag, M. (eds.), *Handbook of Evaluation Research*, Sage, Beverly Hills, Ca, 497-517.
- TAUT, S., 2008, What have we learned about stakeholders involvement in program evaluation, *Studies in Educational Evaluation*, 34, 224-230.

- TEJEDOR, F. J. y otros, 1994, Perspectivas metodológicas actuales de la evaluación de programas en el ámbito educativo, *Revista de Investigación Educativa*, 23, 93-127.
- THE ORAL HISTORY PROJECT TEAM, 2008, The Oral History of Evaluation: The Professional Development of Daniel Stufflebeam, *American Journal of Evaluation*, 29, 4, 555-571.
- THOMPSON, B., 1994, The revised program evaluation standards and their correlation with the evaluation use literature, *Journal of Experimental Education*, 63, 1, 54-82.
- TUKEY, J. W., 1960, Conclusions vs decisions, *Technometrics*, vol. 2, 423-433.
- TUTHILL, D. y ASHTON, P., 1983, Improving educational research through the development of educational paradigms, *Educational Researcher*, 12, 5, 6-14.
- TYLER, R. W., 1950, *Basic principles of curriculum and instruction*, University of Chicago Press, Chicago.
- TYLER, R. W., 1967, Changing concepts of educational evaluation, en Stack, R. E. (comp.), *Perspectives of curriculum evaluation*, AERA Monograph Series Curriculum Evaluation, 1, Rand McNally, Chicago.
- TYLER, R. W. (ed.), 1969, *Educational evaluation: New roles, new means*, University of Chicago Press, Chicago.
- URBAN, J. B. y TROCHIM, W., 2009, The Role of Evaluation in Research-Practice Integration. Working Toward the «Golden Spike», *American Journal of Evaluation*, 30, 4, 538-553.
- WACHTEL, P. L., 1980, Investigation and its discontents. Some constraints on progress in psychological research, *American Psychologist*, 35, 399-408.
- WALKER, J. C. y EVERS, C. W., 1988, The epistemological unity of educational research, en Keeves, J. P. (ed.), *Educational Research, Methodology and Measurement: An International Handbook*, Pergamon Press, Oxford, 28-36.
- WEISS, C. H., 1970, The politization of evaluation research, *Journal of Social Issues*, 26, 4, 57-68.
- WEISS, C. H., 1972, *Evaluating social action programs*, Allyn and Bacon, Boston.

- WEISS, C. H., 1975, Evaluation research in the political context, en Struening, E. L. y Guttentag, M. (eds.), *Handbook of Evaluation Research*, vol. 1, Sage, Beverly Hills, Ca, 13-26.
- WEISS, C. H., 1983, *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción*, Trillas, México.
- WEISS, C. H., 1986, The many meanings of research utilization, en Bulmer, M. y otros (eds.), *Social Science and Social Policy*, Allen and Unwin, London, 31-40.
- WEISS, C. H., 1998, Have we learned anything new about the use of evaluation?, *American Journal of Evaluation*, 19, 1, 21-33.
- WEISS, C. H., 2004, On theory-based evaluation: Winning friends and influencing people, *The Evaluation Exchange*, IX, 4, 1-5.
- WOLF, R. L., 1974, *The citizen as jurist: A new mode of educational evaluation*, Citizen Action in Education, 4.
- WOLF, R. L., 1975, Trial by jury, a new evaluation method, *Phi Delta Kappa*, 57, 185-187.
- WORTHEN, B. R. y SANDERS, J. R., 1973, *Educational Evaluation: Theory and Practice*, Charles A. Jones, Worthington, Ohio.
- WORTHEN, B. R. y SANDERS, J. R., 1991, The changing face of educational evaluation, *Theory into Practice*, XXX, 1, 3-12.
- WRIGHT, C. R., 1967, Evaluation research, en *International Encyclopedia of the Social Sciences*, MacMillan, New York.

# ÍNDICE

PREÁMBULO .....	9
1. INTRODUCCIÓN .....	13
2. EN LA ENCRUCIJADA DE LOS PARADIGMAS DE LA INVESTIGACIÓN .....	15
2.1. Enfoque positivista .....	16
2.2. Enfoque humanístico-interpretativo .....	18
2.3. Enfoque crítico .....	19
2.4. Controversia y convivencia entre paradigmas .....	20
2.5. Características de los paradigmas .....	23
3. EL PARADIGMA DE LA INVESTIGACIÓN EVALUATIVA .....	27
4. LA INVESTIGACIÓN EVALUATIVA, DEFINICIÓN EN UN CONTEXTO DE CAMBIO SOCIAL .....	31
4.1. Algún apunte histórico .....	32
4.2. En interacción con el entorno sociopolítico .....	36
4.3. Precisiones sobre la denominación .....	38
5. LOS GRANDES IMPULSOS DESDE LA EDUCACIÓN .....	43
5.1. La revolución tyleriana .....	43
5.2. La eficacia de los programas y la rendición de cuentas .....	44
5.3. La influencia de Cronbach y Scriven .....	45
5.4. Otras aportaciones de los sesenta .....	49
5.5. La eclosión de modelos evaluativos y la profesionalización .....	51

5.6. La cuarta generación .....	57
5.7. Las recomendaciones de Stufflebeam .....	60
5.8. El auge de la evaluación institucional .....	67
6. SUCINTA SÍNTESIS SOBRE LA CUESTIÓN DE LOS MODE- LOS Y ENFOQUES DE LA INVESTIGACIÓN EVALUATIVA ...	69
7. EL MARCO INVESTIGADOR DEL EVALUADOR .....	81
8. META-EVALUACIÓN Y NORMAS ÉTICAS .....	89
9. UTILIDAD Y USO DE LA EVALUACIÓN .....	99
9.1. Comunicar y diseminar la información adecuadamente ...	104
9.2. Estrategias participativas .....	108
9.3. La toma de decisiones .....	112
9.4. Ética y utilización de la evaluación .....	114
10. REFLEXIÓN PARA FINALIZAR .....	121
REFERENCIAS .....	125

Esta obra se terminó de imprimir  
el 14 de febrero de 2011  
en los talleres gráficos de INO Reproducciones,  
de Zaragoza

