

CLAVES IDENTIFICATIVAS DE LA INVESTIGACIÓN EVALUATIVA: ANÁLISIS DESDE LA PRÁCTICA

Tomás Escudero Escorza

Universidad de Zaragoza

RESUMEN: Se presentan y discuten los elementos definatorios y de identificación más importantes de la investigación evaluativa en las ciencias sociales y en educación en particular, ubicándola en un contexto de cambio y mejora sociales. Desde esta posición y con esta orientación estratégica, se define el alcance y el ámbito de trabajo de la investigación evaluativa, y se presentan y describen las principales visiones y modelos desde los que actúan los investigadores, señalando sus características distintivas y elementos de diferenciación. Un proyecto de evaluación del sistema deportivo riojano sirve como referencia y apoyo para descubrir en una investigación evaluativa real los elementos identificativos antes presentados, así como para ofrecer una reflexión metaevaluativa.

ABSTRACT: In the context of social change and improvement, the most important elements to define and identify evaluation research in social sciences and education are presented and commented. From this position and with this strategic orientation, the scope and space of work of evaluation research are defined, and main visions and models used by researchers are presented and described, with indication of distinctive and differentiation characteristics. An evaluation project of the sportive system of La Rioja is used as a reference and help to discover those identification elements, already presented, in a real evaluation research project, and also to offer a metaevaluative reflection.

No hay duda de que, en los últimos años, la investigación evaluativa se ha convertido en una de las líneas más relevantes de trabajo en el ámbito de las ciencias sociales, en general, y en el de la educación en particular. Cada vez son más los investigadores que trabajan en este campo, y muchos más los que incorporan en sus planteamientos principios, criterios y estrategias de trabajo resaltados por la investigación evaluativa.

En este artículo queremos presentar y reflexionar sobre algunas de las claves, elementos o características que entendemos son más definitorias de lo que hoy se entiende por investigación evaluativa, o evaluación en sentido amplio, tras los innumerables cambios conceptuales y metodológicos que han tenido lugar en este campo desde finales del siglo XIX (Escudero, 2003a). Para ello, nos vamos a servir de un trabajo inédito, realizado hace unos años, de evaluación del sistema deportivo, incluido el escolar, de la Comunidad Autónoma de La Rioja, porque contiene elementos de reflexión práctica abundantes sobre gran parte de las claves de la investigación evaluativa que queremos analizar.

Pudiera extrañar que hayamos tomado como apoyo un ejemplo no exclusivamente del ámbito educativo, pero lo hacemos así para resaltar que la investigación evaluativa, aunque haya tenido un desarrollo especialísimo en educación, se trata de un fenómeno de incidencia más amplia y trascendente, sobre todo, para todo el conjunto de las ciencias sociales.

Pero no debemos olvidar que este trabajo no centra su interés primordial sobre el estudio del sistema deportivo riojano, que sólo es un soporte, entendemos que adecuado, para analizar algunas claves de la investigación evaluativa, que sí es nuestro principal centro de interés.

1. EN CONTEXTO DE CAMBIO SOCIAL

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 y la investigación social aplicadas 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, está en su propia raíz paradigmática (Carr y Kemmis, 1983; Morin, 1985; De Miguel, 1988; Nisbet, 1988; Fishman, 1991; Worthen y otros, 1997) y, además, es un elemento reconocible en los trabajos o estudios que aparecen en este campo (Cracknell, 2000; Fernández-Ballesteros, 2001; Escudero, 2003b).

Para perfilar este marco de actuación, nos conviene recordar aquí los elementos definitorios del que Fishman ha denominado como *paradigma pragmático*, híbrido entre el experimental y el hermenéutico, en el que la investigación evaluativa se sustenta. 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 trabaja con *grandes volúmenes de datos*.
- i) Se enfatiza la *predictibilidad* de la conducta.
- j) Se busca la *mejora* de los *programas sociales*.
- k) Se informa a los responsables de *tomar decisiones* sobre programas y prácticas.

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 marca 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, asociaciones científicas, cursos y textos múltiples (Berk y Rossi, 1990; Clarke y Dawson, 1999). Otros autores como Gredler (1996), ubicarían el comienzo de la investigación evaluativa en 1932, con el comienzo del 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.

Un sello de la investigación evaluativa moderna es la aproximación científica para evaluar el rendimiento de los programas Wright (1967). En este sentido, la investigación evaluativa se equipara a otras formas de investigación social, pero la especificidad viene de los objetivos y de las condiciones de realización. Para entendernos más claramente, digamos que se consideran programas sociales los de epidemiología y salud (Susser, 1975), pero no estudios biológicos de laboratorio con animales, por ejemplo.

En los años cincuenta y primeros sesenta se hablaba principalmente de evaluación de programas sociales y, de hecho, una denominación habitual de este campo de trabajo ha sido durante bastante tiempo la de *evaluación de programas*, pero finalmente se ha impuesto el término de *investigación evaluativa*, porque recoge mejor la amplitud del campo, mientras que la evaluación de programas se limita a la investigación evaluativa directamente relacionada con programas sociales (Owen y Rogers, 1999).

La evaluación existe para facilitar la toma de decisiones inteligente (Edwards y otros, 1975), de forma que si no se cumple esto en el contraste entre programas que compiten, pierde su especificidad. *La investigación evaluativa, entonces, es un tipo especial de investigación social en relación con los programas, planes e instituciones sociales y con la toma de decisiones*. En este contexto, es lógico, tal como se hace hoy (McClintock, 2003; Calderon, 2004), definir al evaluador como un *agente del cambio*.

Hace unos años, Kish (1987) señalaba las principales notas distintivas de la investigación evaluativa, tomando como referencia de comparación la investigación social en general y destacaba las siguientes:

- a) La investigación evaluativa tiene que ver con las *políticas 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 evaluador 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 falle, 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 que deben ser balanceados con los positivos. También hay que buscar la detección de efectos no esperados (serendipity). Hay que analizar la eficacia en relación a los costos, *eficiencia*, y desde múltiples perspectivas.
- f) La difusión de la investigación evaluativa sigue unos cauces diferentes que 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.
- g) 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 todavía siguen existiendo problemas. [Esto nos trae a la mente un punto de actual debate en nuestro país, como es el problema de enmarque legal, ubicación y entronque institucional de la Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) y de las diferentes Agencias Autonómicas].

Carol H. Weiss, uno de los máximos referentes en el campo de la investigación evaluativa, sobre todo en los primeros setenta (Weiss, 1970, 1972, 1973, 1975 y 1998), no sólo 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 impacto político. 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. Finalmente y en tercer lugar, por algo que según Weiss es menos reconocido, esto es, porque la propia 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 métodos, la utilidad de las estrategias de reforma propuestas, e incluso la influencia lógica del científico social en la política.

Este último argumento de Weiss entendemos que puede matizarse, pero sí que creemos que ni el evaluador se mueve en un *contexto neutro*, ni el objeto que analiza es inmune a presiones que se escapan 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, existiendo ciertas estrategias que se suelen utilizar en este sentido (Escudero, 1980).

Inciendo también en esta línea, Berk y Rossi (1990) señalan al menos tres conceptos claves en la investigación evaluativa, esto es, 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.

El evaluador está obligado a respetar los derechos de todos los individuos implicados o afectados por el estudio y también a las obligaciones contraídas con los promotores del mismo y con los responsables de la toma de decisiones. Es por ello que, 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.

Stake (1998) incide también 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, según su opinión, atienden básicamente los intereses de los contratantes de las evaluaciones, descuidando otros intereses y la propia calidad de los programas. En consonancia con esta misma línea de pensamiento, 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 sus voces en la evaluación; se necesita un paso más. Sus voces proporcionan más credibilidad a la evaluación, pero la transformación sólo puede ocurrir realmente si su información es usada para fundamentar programas sociales que intenten atajar las desigualdades existentes.

La mayoría de los estudios meta-evaluativos sobre ética y evaluación se centran en el evaluador y, más en concreto, en sus principios y en su práctica (Christie y Alkin, 1999). En otras palabras, se resalta el peligro de las malas evaluaciones (*mis-evaluations*), mal orientadas, sesgadas, incompletas o débiles metodológicamente. Pero estos autores 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.

Las reflexiones anteriores nos enmarcan de manera clara un *tipo de investigación aplicada, que incide sobre objetos sociales, programas, centros, agentes, etc., que analiza y juzga su calidad estática y dinámica según criterios científicos rigurosos múltiples, con la obligación de sugerir, a las diversas audiencias implicadas, acciones alternativas sobre los mismos para diferentes propósitos como utilización, mejora, acreditación, fiscalización, reforma, etc.* (Escudero, 1996).

Esta definición que acabamos de ofrecer de la investigación evaluativa es muy similar a las múltiples que nos encontramos en la literatura especializada bajo esta misma denominación (Borg y Gall, 1979; McMillan y Schumacher, 1984; Arnal y otros, 1992) o bajo la denominación de evaluación de programas (De la Orden, 1985; Colás y Rebollo, 1993), limitando a los programas los citados objetos sociales. De la Orden, 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 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 educativos posibles. Para nosotros, desde luego, la investigación evaluativa es mucho más que la evaluación de programas en sentido estricto, aunque reconozcamos a esta en el ori-

gen y en el núcleo de aquélla. Esta posición es la defendida por un autor de la relevancia de Scriven.

El que fue primer presidente de la «American Evaluation Association» y, sin duda, el principal responsable de nuestros grandes debates sobre *evaluación formativa vs sumativa* (Scriven, 1967), nos ofrece un interesantísimo ensayo reconceptualizando lo que 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 denominadas por Scriven 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. Como ya hemos indicado anteriormente, Nevo (1989), en un ensayo anterior sobre la conceptualización de la evaluación, también señala la multiplicidad de objetos de la evaluación.

El que esta disciplina general haya tardado en asentarse definitivamente, a pesar del largo uso de la denominación, es, según Scriven, el uso 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.

2. VISIONES Y MODELOS PARA LA INVESTIGACIÓN EVALUATIVA

Como respuesta a los problemas que acabamos de apuntar, Scriven (1994) nos ofrece una clasificación de los modelos de evaluación, previamente a introducir su modelo transdisciplinar que más adelante comentamos. Este autor identifica seis visiones o enfoques alternativos en la fase «explosiva» de los modelos que tuvo lugar en los años setenta (Guba y Lincoln, 1982; Mateo, 1986; Escudero, 2003a), además de algunas más que denomina «exóticas» y que se mueven entre los modelos de jurisprudencia y de experto. A continuación presentamos 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 le 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 extensamente elaborada en el modelo CIPP (Stufflebeam y otros, 1971).

Según el planteamiento tyleriano, las decisiones acerca de un programa deben basarse en el grado de coincidencia entre los objetivos y los resultados. El cambio de los alumnos, habitualmente el objetivo perseguido, es el criterio de evaluación.

A diferencia de Tyler, Stufflebeam ofrece una perspectiva más amplia de los contenidos a evaluar, a 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. Stufflebeam ha seguido desarrollando su perspectiva desde que desarrolló el CIPP. Sin embargo, uno de sus colaboradores, Guba, ha seguido posteriormente por una dirección diferente, encabezando la denominada cuarta generación de la evaluación (Guba y Lincoln, 1989).

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 a la evaluación como un proceso factual de recogida y generación de información al servicio del que toma las decisiones, pero es éste el que tiene que tomar 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. Este autor (Alkin, 1991) revisó dos décadas después su planteamiento inicial pero 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.

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 junto con la anterior han sido el camino que ha permitido su integración sin problemas en el «carro» de la investigación evaluativa. De hecho, uno de los textos más utilizados de evaluación (Rossi y Freeman, 1993), está escrito bajo esta perspectiva. Pero la forma más simple de la visión relativista (visión C) fue desarrollada dentro del «modelo de discrepancia» de evaluación de Malcolm Provus (1971). Las discrepancias son las divergencias con la secuencia de tareas proyectadas y la temporaliza-

ción prevista. Este modelo es muy cercano al control de programas en sentido convencional; es una especie de simulación de una evaluación.

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 de la educación, la educación comparada y la administración educativa.

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 evaluador informa de lo que ve sin intentar emitir afirmaciones valorativas o inferir conclusiones evaluativas, ni siguiera 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 como *visión de la descripción sólida, fuerte*, para evitar el término *rica*, que parece más evaluativa.

Stake, en su primera etapa, es tayleriano 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 mitad de los setenta, Stake se aleja de la tradición tayleriana de preocupación por los objetivos y revisa su método de evaluación hacia un planteamiento que él califica como «*respondente*» (Stake, 1975 y 1975a), asumiendo que los objetivos del programa pueden modificarse sobre la marcha, con la finalidad de ofrecer una visión del mismo completa y holística 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.

De manera gráfica, Stake (1975a) propone las fases del método a modo de las horas de un reloj, poniendo la primera en las doce horas y tomando el sentido de las agujas del reloj. Estas fases son las siguientes: 1) Hablar con los clientes, responsables y audiencias, 2) Alcance del programa, 3) Panorama de actividades, 4) Propósitos e intereses, 5) Cuestiones y problemas, 6) Datos para investigar los problemas, 7) Observadores, jueces e instrumentos, 8) Antecedentes, transacciones y resultados, 9) Desarrollo de temas, descripciones y estudio de casos, 10) Validación (confirmación), 11) Esquema para la audiencia y 12) Reunión de informes formales. El evaluador puede seguir las fases también en sentido contrario del reloj o en cualquier otro orden.

En el método respondiente, 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 (Stecher 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 más de dos 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. De hecho, Cronbach entiende a la evaluación de programas como un instrumento importante de desarrollo social y comunitario (Escudero, 2004). Sus posiciones quedaban claramente establecidas en noventa y cinco tesis que han tenido una enorme difusión entre los evaluadores y los usuarios de la evaluación.

En cuanto a los contenidos de la evaluación, Cronbach (1982) propone que se planifiquen y controlen los siguientes elementos:

- *Unidades* (U) que son sometidas a evaluación, individuos o grupos participantes.
- *Tratamiento* (T) de la evaluación.
- *Operaciones* (O) que lleva a cabo el evaluador para la recogida y análisis de datos, así como para la elaboración de conclusiones.
- *Contexto* en el que tiene lugar el programa y su evaluación.

En una investigación evaluativa concreta se pueden dar varias unidades, varios tratamientos y varias operaciones, en definitiva varios (uto), dentro de un universo UTO de situaciones admisibles.

Ernie House (1989), un prestigioso teórico y un práctico de la evaluación bastante independiente de las diversas corrientes, también marcó el entronque social de los programas, pero se distinguía sobre todo por enfatizar las dimensiones más éticas y argumentales de la evaluación, quizás en contraposición a 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, mantenida por Guba y Lincoln (1989) y seguida por muchos evaluadores americanos y británicos. Ya hemos visto anteriormente que 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 le ocurre a todo trabajo de análisis filosófico como el suyo. Scriven apunta que el propio Guba ha sido siempre consciente de las potenciales «contradicciones» de su posición.

De esta revisión 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 (1975) proponen un método de *contraposición* o discusión que sobre un programa llevan a cabo dos grupos

de evaluadores, partidarios y adversarios, para proporcionar información pertinente a quienes toman decisiones. Eisner (1975) plantea la evaluación en términos similares al proceso de crítica artística.

El propio Scriven (1973) proponía hace años 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 como *evaluación sin metas*. El evaluador determina el valor o mérito del programa para informar a los usuarios; se trata algo así como de un intermediario informativo (Scriven, 1980).

La evaluación iluminativa (Parlett y Hamilton, 1977) tiene un enfoque holístico, descriptivo e interpretativo, con la pretensión de iluminar sobre un complejo rango de cuestiones que se dan de manera interactiva (Fernández, 1991). *La evaluación democrática* de McDonald (1976), también denominada holística, supone la participación colaborativa de los implicados, siendo el contraste de opiniones 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, que 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 modelo que denomina *transdisciplinar* y que él califica como significativamente distinta de la aludida visión A y radicalmente diferente de las restantes.

En el *modelo 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 le ocurre a disciplinas como la estadística y la medición. 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*.

Esta visión es *objetivista* como la visión A y defiende que el evaluador determine el mérito o el valor del programa, del personal o de los 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 perseguir las falacias argumentales de la doctrina libre de valores (Evaluation Thesaurus, 1991).

En segundo lugar, la visión transdisciplinar se orienta hacia el *consumidor*, más que hacia el gestor o intermediario. No se trata de una orientación exclusiva hacia el consumidor, pero sí la consideración primera del consumidor como justificación del programa, y que el bien común es 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 sólo ve legitimidad en la emisión de conclusiones evaluativas por parte del investigador, sino que ve necesidad de 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 práctica. Desde esta perspectiva, la investigación evaluativa es mucho más que la evaluación de programas e incide en procesos, instituciones y otros muchos objetos. De manera más detallada, esta visión generalizada significa que:

- a) Los campos distintivos de aplicación de la disciplina son los programas, el personal, los rendimientos, los productos, los proyectos, la gestión, y la metaevaluación de todo ello.
- b) Las investigaciones evaluativas inciden en todo tipo de disciplinas y en las prácticas que resultan de ellas.
- c) Las investigaciones evaluativas se mueven desde niveles muy prácticos hasta el nivel conceptual.
- d) Los distintos campos de la investigación evaluativa tienen muchos niveles de interconexión y solapamiento. La evaluación de programas, de personal, de centros, etc., tienen muchos puntos en común.

El cuarto elemento distintivo de la visión transdisciplinar de la evaluación es que se trata de una visión *técnica*. La evaluación no sólo necesita el apoyo técnico de otras muchas disciplinas, sino que, además, tiene su propia metodología que es sustancial para la evaluación. Por otra parte, están la lógica de la síntesis de resultados, las consecuencias, etc., así como 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 visión transdisciplinar de la investigación evaluativa coincide en gran medida con los planteamientos que de la misma hemos presentado al principio del artículo. 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 tienen 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 entendemos que esta moderna visión de Scriven es sólida y coherente y que se acerca bastante a la que sería nuestra personal posición a la hora de conceptualizar la investigación evaluativa. Quizás nuestra única discrepancia de cierta entidad con Scriven es el excesivo énfasis relativo que da a la orientación al cliente, al usuario en sentido estricto. Nosotros entendemos esto 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 creemos que la investigación evaluativa debe tener una orientación prioritaria algo más plural que la defendida por este autor.

3. LA EVALUACIÓN DEL SISTEMA DEPORTIVO DE LA RIOJA COMO UN EJEMPLO

Tal como hemos anunciado previamente, el estudio en el que apoyamos nuestras reflexiones prácticas es una investigación evaluativa del sistema deportivo riojano, llevada a cabo a mediados de los noventa, con un amplio rango de objetivos de elementos de análisis, de planteamientos metodológicos y de procedimientos

técnicos. Fue encargado por la Consejería de Cultura, Deportes y Juventud del Gobierno de La Rioja, a través de su Dirección General de Deportes, pero realizado de manera autónoma por el equipo investigador.

El trabajo es una evaluación diagnóstica de la situación del deporte riojano en muchas facetas, pero también tiene orientación prospectiva. Se evaluaron las necesidades y se buscaron líneas de mejora alternativas. En algunos momentos se describió la situación de manera exhaustiva y se trabajó como si el objetivo fuera la elaboración de un censo determinado, aunque nunca fue éste un producto obligado para el proyecto. Otras veces se realizaron sondeos en sentido estricto y algunas otras se buscaron tendencias indicativas y sugerencias cualificadas para proponer acciones de intervención. Consecuentemente con lo dicho, el estudio se encuadra en la investigación evaluativa orientada hacia la toma de decisiones en el terreno de la planificación e intervención en materia de política deportiva.

El marco referencial desde el que se fundamentó el proyecto es el de que el sistema deportivo de un territorio es el conjunto estructurado de instituciones, servicios, instalaciones, esquemas organizativos, deportistas cualificados, dirigentes, técnicos, jueces,..., practicantes en general, siendo el fenómeno deportivo, la competición, la actividad, la práctica, la observación,..., el producto natural de dicho sistema.

Dentro de esta perspectiva definitoria, el objetivo central del estudio era:

- *El diagnóstico preciso de la realidad del sistema deportivo en La Rioja, en lo referente a sus componentes principales (deportistas, practicantes en general, instalaciones, equipamientos, infraestructuras, asociaciones, clubes, federaciones, responsables, técnicos, jueces, personal de apoyo,...) y la búsqueda de acciones alternativas de mejora y desarrollo.*

Para alcanzar este propósito, se establecieron como objetivos específicos los siguientes:

- a) La descripción completa del sistema deportivo riojano en el momento en todas sus facetas: personal, organizaciones, recursos y actividad.
- b) El análisis de necesidades y establecimiento de prioridades de acción para los distintos componentes del sistema deportivo (instalaciones, técnicos, recursos, formación, apoyos, etc.).
- c) La descripción y valoración diferencial de los distintos ámbitos de actuación y responsabilidad del sistema deportivo (autonómico, municipal, escolar, federativo y asociativo).
- d) La evaluación del sistema deportivo por zonas geográficas, por modalidades y por niveles deportivos.
- e) El análisis de las relaciones institucionales y estamentales dentro del sistema deportivo.
- f) La búsqueda de alternativas de desarrollo en materia de instalaciones, recursos, personal, etc.
- g) La valoración del sistema y de sus líneas de mejora desde las distintas perspectivas institucionales, de los expertos, de los deportistas y de los practicantes en general.

Para el cumplimiento de los anteriores objetivos, se establecieron unos ámbitos diferenciados de análisis y, dentro de ellos, se estudiaron en profundidad las variables consideradas más relevantes. Estos ámbitos eran los siguientes:

- A. El deporte autonómico (la Dirección General de Deportes).
- B. El deporte municipal (los Ayuntamientos).
- C. El deporte escolar (los Centros educativos).
- D. El deporte federativo (las Federaciones territoriales).
- E. El deporte asociativo (los Clubes y las Asociaciones deportivas).
- F. Las instalaciones deportivas.
- G. La visión de los dirigentes, expertos y responsables deportivos.
- H. La visión de los deportistas federados.
- I. La visión del deporte de los riojanos en general.
- J. Algunos elementos de contraste externo.

Estos diez ámbitos se recomponen en tres bloques perfectamente diferenciados. El primer bloque lo forman los cinco primeros ámbitos de análisis, esto es, los deportes autonómico, municipal, escolar, federativo y asociativo, cuyo estudio fue llevado a cabo de forma similar y paralela para todos ellos, teniendo presente como es lógico las peculiaridades de cada caso. De esta manera, se pueden establecer valoraciones comparativas entre los distintos ámbitos.

El segundo bloque lo constituye un ámbito específico, el de las instalaciones deportivas, elemento central de la investigación por múltiples y variadas razones.

Finalmente, el tercer bloque, de carácter más complementario, dentro del proyecto global, pero de gran interés como fuente de ideas, recoge el estudio de las opiniones de tres sectores importantísimos en el deporte que no necesariamente han sido consultados en el primer bloque: los dirigentes y concededores del deporte, los deportistas federados y los ciudadanos en general. Además, se incluyen algunos datos de comparación de la situación riojana con la de otros contextos.

Es obvio decir que en cada uno de estos bloques fueron investigadas decenas de variables distintas que no tiene sentido reproducir en este momento. Basta decir que dichas variables, con sus matices correspondientes, debidamente codificadas, antes o después de recogida la información, fueron analizadas a lo largo del proyecto, al menos desde una perspectiva. En muchos casos la perspectiva fue múltiple, apareciendo elementos comparativos y de confirmación.

Aparte de las fuentes documentales y bibliografía utilizada a lo largo del trabajo, éste se fundamentó preferentemente en la recogida de información primaria de los principales protagonistas institucionales y particulares del sistema deportivo.

El vehículo de comunicación del equipo investigador con las diversas fuentes fue diverso. Con la Dirección General de Deportes la comunicación fue directa, de primera mano. A los Ayuntamientos, Centros educativos, Federaciones y Clubes y Asociaciones se les aplicó un cuestionario específico para cada caso, enviado y devuelto por correo. Las instalaciones deportivas fueron visitadas una a una por un colaborador del equipo investigador quien, apoyándose en las correspondientes fichas de instalaciones (general e individual), recogieron toda la información requerida. Los

dirigentes y responsables deportivos fueron entrevistados con un protocolo de entrevista dirigido, pero abierto, rellenado por su cuenta y devuelto de manera anónima. Los deportistas federados fueron encuestados por correo. Finalmente, los ciudadanos en general fueron entrevistados siguiendo un protocolo de cuestiones cerradas.

Los correspondientes análisis de los datos cuantitativos y de las informaciones cualitativas recogidos fueron siempre llevados a cabo, y debidamente contextualizados, para el conjunto de toda la Comunidad Autónoma de La Rioja y para cada una de las diez zonas deportivas en la que está dividida a efectos de gestión. En determinados casos, apoyados en paquetes informáticos convencionales, se llevaron a cabo análisis interactivos por edad, sexo, modalidad deportiva, nivel competitivo y categoría deportiva.

4. RECONOCIMIENTO DE LA INVESTIGACIÓN EVALUATIVA EN EL PROYECTO ANTERIOR

En este proyecto sobre el deporte riojano que acabamos de presentar de manera sucinta, podemos reconocer un buen número de elementos identificativos de la investigación evaluativa a los que de manera directa o implícita hemos hecho referencia a lo largo de este trabajo. En tal sentido, creemos que el estudio es un buen ejemplo, en cuanto a potencial pedagógico, para mostrar algunas claves de especificidad de este tipo de investigación.

En el “Análisis del sistema deportivo de La Rioja” nos encontramos con un conjunto de características que son típicas, aunque no necesariamente exclusivas, de la investigación evaluativa. Entre estas características destacan las siguientes:

- Investigación por contrato.
- Respuesta a una solicitud.
- Equipo amplio de investigación.
- Contexto de solución de problemas.
- Ideografía (Descripción de caso particular).
- Contexto transdisciplinar.
- Trabajo en situación natural.
- Desarrollo de metodología “ad hoc”.
- Varias audiencias implicadas.
- Datos de todo tipo. Cuantitativos y cualitativos.
- Apoyo básico en la observación.
- Grandes volúmenes de información.
- Presión política envolvente.
- Amplia implicación social.
- Plazos de realización.
- Cooperación entre “responsables” (solicitantes) e investigadores.
- Conocimiento de la realidad y diagnóstico de necesidades.
- Concentración de esfuerzos en necesidades prioritarias.
- Sugerencias para toma de decisiones y políticas de mejora.

Si nos centramos ahora en la orientación preferente del trabajo, esto es, en aquello que mejor sintetiza su naturaleza como estudio, también detectamos algunos elementos específicos de un tipo de investigación evaluativa centrada en el *análisis de necesidades* (Moroney, 1977; Kamis, 1981; Stufflebeam, 1985). En concreto nos referimos a los siguientes:

- Evaluación de necesidades con orientación diagnóstica para la planificación (needs assessment).
- Búsqueda de necesidades con apoyo de distintas perspectivas.
 - De la discrepancia (logros vs deseos).
 - Democrática (deseos mayoritarios).
 - Analítica (mejora posible según datos disponibles).
 - Diagnóstica (algo ausente, deficiente o perjudicial).
- Atención a necesidades diversas.
 - Normativas (criterios, indicadores, expertos,...).
 - Percibidas (por individuos).
 - Expresadas (demandas).
 - Relativas (entre situaciones, grupos,... distintos).
- Detección e identificación de necesidades con distintas estrategias.
 - Directas (archivos, estadísticas).
 - Opiniones (percepción individuos).
 - Funcionamiento de servicios (inferir de observación).
 - Grupos asociativos (información y expresiones diversas).

En cuanto a planteamientos metodológicos, la postura desde la que se abordó la evaluación del sistema deportivo riojano fue claramente ecléctica, ante las posibilidades de apoyo en diferentes enfoques modélicos. Esta posición como evaluador no es el reflejo de falta de compromiso o de defensa ante la confrontación entre paradigmas, sino del convencimiento de que es la lógica y la científicamente más sólida para que la investigación evaluativa sea rigurosa y coherente con sus principios definitorios.

Retomando la terminología que hemos introducido y utilizado en apartados anteriores, diremos que el apoyo fundamental para diseñar metodológicamente el proyecto estuvo en el modelo CIPP (Stufflebeam) y en la visión fuerte (visión A) hacia la toma de decisiones. Esto se refleja en el énfasis de algunos elementos tales como:

- * El análisis del contexto, de los "inputs" (recursos vs necesidades), del proceso (dinámica, actividades) y productos (impactos).
- * La prioridad en la mejora.
- * El riesgo tomado como evaluadores al emitir juicios sobre posibles decisiones de intervención.
- * El apoyo en el sentido común.

Algunas veces preveleía la visión débil (visión B) hacia la toma de decisiones (Alkin), pues se proporcionaba mucha información cualitativa y cuantitativa, para distintas audiencias, pero sin propuestas de intervención.

Otro apoyo decisivo fue la visión de la descripción fértil, rica y completa (visión D), que recoge la posición respondiente de Stake. Esto es evidente en los análisis de confluencias y diferencias de opinión de los distintos implicados en el deporte riojano, tan decisivos en el desarrollo del estudio.

La visión E (Cronbach) está en el sustrato de todo el proyecto, pues parte del principio de que el deporte es algo de gran importancia social y de que el objetivo principal de su evaluación es intervenir para mejorarlo. Además, esto se intenta, introduciendo en el estudio y en su desarrollo a los distintos sectores, colectivos y responsables implicados en el hecho deportivo (visión F-Guba y Lincoln) y buscando esquemas de adaptación a sus diversos intereses y necesidades.

La visión transdisciplinar (Scriven) es reconocible en muchos momentos a lo largo del proyecto, no sólo por la diversidad temática abordada y por el apoyo en un buen número de disciplinas, sino por la postura objetivista del equipo investigador, emitiendo juicios de valor sobre lo encontrado, por los análisis generalizados, por las técnicas metodológicas y, en general, por una aceptación clara del pragmatismo como paradigma de trabajo.

En los múltiples y extensos análisis de la información recogida se limitó siempre la utilización de índices cuantitativos a aquellos que fueran interpretables y útiles para las diversas audiencias. Así mismo, se enfatizó la interpretación de tendencias más que la producción de índices múltiples.

A la hora de establecer juicios de valor, el evaluador necesita referentes, que no siempre son fáciles de encontrar. En muchos casos la salida lógica es el análisis comparado con otros sistemas similares. En el caso que nos ocupa, este problema era especialmente grave, pues las referencias publicadas eran muy escasas y, además, con distintos enfoques conceptuales, a su vez muy diversos entre sí. En todo caso, la búsqueda de algunos referentes en el análisis comparado y en publicaciones sobre el deporte y su práctica en contextos nacionales e internacionales, fue un ejercicio típico de la investigación evaluativa del que, en este caso, se obtuvieron algunos logros interesantes para valorar el sistema riojano.

5. REFLEXIONES METAEVALUATIVAS

La metaevaluación es algo que forma parte de la esencia de cualquier proyecto de investigación evaluativa. No tendría legitimidad el trabajo de un equipo investigador que se dedica a la evaluación de programas sociales, que no evaluara su propio programa de evaluación, máxime cuando se realiza en contexto real y dentro de unos condicionantes como los antes apuntados, que establecen algunos límites en cuanto al rigor y al alcance de la investigación.

En el estudio del deporte riojano, sin duda nos encontramos con algunas limitaciones de alcance y significación en sus diferentes partes que, por su carácter ejemplificador debemos resaltar.

El contexto de caso particular, con unas especificidades muy notables desde el punto de vista geográfico y administrativo, así como de las demandas y necesidades que había que atender, no cabe duda que condicionó el enfoque de investigación hacia un análisis más sociológico que económico o de eficiencia en la utilización de recursos.

El estudio planteaba una evaluación sin objetivos definidos previamente de manera explícita, lo que imposibilitaba todo análisis riguroso de la eficacia del sistema deportivo y de sus componentes y de su eficiencia en la utilización de recursos. No había otro remedio que limitarse a evaluar la productividad de recursos e infraestructuras y la pertinencia con relación a objetivos sociopolíticos amplios o incluso generales.

El estudio se hizo con referentes metodológicos previos escasos y sin criterios de contraste reconocidos. Esta situación no es infrecuente para el investigador evaluativo, pues es habitual que reciba el encargo de evaluar programas o instituciones muy específicos, que no han sido evaluados con anterioridad, o que actúan en contextos muy particulares, condicionando no sólo su incidencia, sino su propia realidad.

Ante la dificultad evidente del análisis comparado por las carencias citadas, es muy importante extremar la prudencia científica y apoyarse en la triangulación de fuentes, a la búsqueda de confluencia de criterios. En este estudio, por ejemplo, se evitó en todo momento hablar de indicadores, se trató siempre con variables, buscando espacios de confluencia en las valoraciones de expertos, responsables políticos, dirigentes, técnicos, deportistas y ciudadanos en general.

BIBLIOGRAFÍA

- ALKIN, M., 1969, Evaluation theory development, *Evaluation Comment*, 2 (1), 2-7.
- 1991, Evaluation theory development: II, en McLaughlin, M. y Phillips (Eds.), *Evaluation and education at quarter century*, NSSE/ University of Chicago, Chicago, 91-112.
- ARNAL, J. y otros, 1992, *Investigación educativa. Fundamentos y metodología*, Labor, Barcelona.
- BERK, A. y ROSSI, P. H., 1990, *Thinking about program evaluation*, Sage Publications, Newbury Park, Ca.
- BORG, W. R. y GALL, M. D., 1979, *Educational research: An introduction*, Longman, New York.
- CALDERON, A. J., 2004, Institutional Research at RMIT. A case study, Ponencia presentada en el *26th EAIR Forum*, Barcelona, 5-8 de septiembre.
- CARR, W. y KEMMIS, S., 1983, *Becoming critical: Knowing through action research*, Deakin University, Victoria, Australia.
- CHRISTIE, Ch. A. y ALKIN, M. C., 1999, Further reflections on evaluation misutilization, *Studies in Educational Evaluation*, 25 (1), 1-10.
- CLARKE, A. y DAWSON, R., 1999, *Evaluation Research*, Sage Publications, London.
- COLÁS, P. y REBOLLO, M. A., 1993, *Evaluación de programas. Una guía práctica*, Kronos, Sevilla.
- CRACKNELL, B. E., 2000, *Evaluating Development Aid*, Sage Publications, New Delhi.

- CRONBACH, L. J., 1982, *Designing evaluations of educational and social programs*, Jossey-Bass, Chicago.
- y otros, 1980, *Towards reform in program evaluation: Aims, methods and institutional arrangements*, Jossey-Bass, San Francisco.
- 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., 1988, Paradigmas de la investigación educativa española, en Dendaluce, I. (Coord.), *Aspectos metodológicos de la investigación educativa*, Narcea, Madrid, 60-77.
- EDWARDS, W. y otros, 1975, A decision-theoretic approach to evaluation research, en Struening, E. L. y Guttentag, M. (eds.), *Handbook of Evaluation Research*, Sage Publications, Beverly Hills, Ca., 139-181.
- EISNER, E. W. 1975, *The perceptive eye: Toward the reformation of educational evaluation*, Stanford Evaluation Consortium, Stanford, Ca.
- ESCUADERO, T., 1980, *¿Se pueden evaluar los centros educativos y sus profesores?*, Educación Abierta, 10, ICE-Universidad de Zaragoza.
- 1996, *Proyecto docente e investigador*, Universidad de Zaragoza.
- 2003a, Desde los tests hasta la investigación evaluativa actual. Un siglo, el XX, de intenso desarrollo de la evaluación en educación, *RELIEVE- Revista Electrónica de Investigación y Evaluación Educativa*, 9 (1). Disponible en: http://www.uv.es/RELIEVE/v9n1/RELIEVEv9n1_1.htm. [Última consulta: 22/03/2006]
- 2003b, Orientar con programas: su evaluación en clave social, Conferencia inaugural del Master de "Avaliación e Orientación de Competencias e Cualificaciones Profesionais", Universidad de Santiago de Compostela.
- 2004, Las tesis de Cronbach sobre la evaluación de programas: El pensamiento de un maestro, en Soler, J. L. (Coord.), *Orientación y Tutoría*, Mira Editores, Zaragoza, 149-158.
- y otros, 1994, *Análisis del Sistema Deportivo de la Rioja*, Consejería de Cultura, Deportes y Juventud del Gobierno de La Rioja. Memoria final inédita.
- EVALUATION THESAURUS, 1991, Sage, Newbury Park, Ca.
- FERNÁNDEZ, J., 1991, La evaluación de la calidad docente, en Medina, A. (coord.), *Teoría y métodos de evaluación*, Cincel, Madrid.
- FERNÁNDEZ-BALLESTEROS, R. (ed.), 2001, *Evaluación de programas: una guía práctica en ámbitos sociales, educativos y de salud*, Síntesis, Madrid.
- FISHMAN, D. B., 1991, An introduction to the experimental versus the pragmatic paradigm in evaluation, *Evaluation and Program Planning*, 14 (4), 353-363.
- GREDLER, M. E., 1996, *Program evaluation*, Merrill/Prentice-Hall, Englewood Cliffs, N.J.
- GUBA, E. G. y LINCOLN, Y. S., 1982, *Effective evaluation*, Jossey Bass Publishers, San Francisco.
- 1989, *Fourth Generation Evaluation*, Sage Publications, Newbury Park, Ca.
- HERZOG, E., 1959, *Some Guidelines for Evaluative Research*, U.S. Department of Health, Education, and Welfare, Washington, D.C.

- HOUSE, E. R., 1973, Politics as usual, en House, E.R. (ed.), *School Evaluation. The Politics and the Process*, McCutchan Publishing Corporation, Berkeley, 329-331.
- 1989, *Evaluating with validity*, Sage, Newbury Park, Ca.
- JOINT COMMITTEE on Standards for Educational Evaluation, 1994, *The program evaluation standards*, Sage, Thousand Oaks, CA.
- KAMIS, E., 1981, Sound targeted comparison: Assessing the needs of planning services for deinstitutionalized clients, en Rutman, I. D. (ed.), *Planing for deinstitutionalization: A review of principles, methods and applications*, Human Services Monograph Series nº 28, Project SHARE, Department of Health and Human Services, Washington.
- KISH, L., 1987, *Statistical Design for Research*, John Wiley & Sons, New York.
- MATEO, J., 1986, *Proyecto docente e investigador*, Universidad de Barcelona.
- MCCLINTOCK, C., 2003, Commentary: The evaluator as scholar/practitioner/change agent, *American Journal of Evaluation*, 24 (1), 91-96.
- MCDONALD, B., 1976, Evaluation and the control of education, en Tawney, D. (Ed.), *Curriculum evaluation today: Trends and Implications*, MacMillan, London.
- MCMILLAN, J. H. y SCHUMACHER, S., 1984, *Research in education*, Little, Brown & Co., Canada.
- MERTENS, D. M., 1999, Inclusive evaluation: Implications of transformative theory for evaluation, *American Journal of Evaluation*, 20 (1), 1-14.
- MORIN, A., 1985, Criteres de scientificité de la recherche-action, *Revue des Sciences de l' Education*, 11(1), 31-49.
- MORONEY, R. M., 1977, Needs assessment for Human Services, en Anderson, W. F. y otros (Eds.), *Managing Human Services*, International City Management Association, Washington.
- 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., 1988, Policy-oriented research, en Keeves, J.P. (Ed.), *Educational research, methodology and measurement: An international handbook*, Pergamon Press, Oxford, 139-146.
- OWENS, T. R., 1973, Educational evaluation by adversary proceedings, en House, E. R. (comp.), *School Evaluation: The Politics and Process*, McCutchan, Berkeley.
- OWEN, J. M. y ROGERS, P. J., 1999, *Program evaluation. Forms and approaches*, Sage Publications, London.
- 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.
- PROVUS, M., 1971, *Discrepancy evaluation. For educational program improvement and assessment*, McCutchan Publishing Co., Berkeley, Ca.
- ROSSI, P. H. y FREEMAN, H., 1993, *Evaluation: A Systematic Approach*, Sage, Beverly Hills, Ca.
- SCRIVEN, M., 1967, The methodology of evaluation, en *Perspectives of Curriculum Evaluation*, (AERA Monograph 1), Rand McNally and Company, Chicago, 39-83.

- 1973, Goal-free evaluation, en House, E. R. (Ed.), *School evaluation: The politics and process*, McCutchan, Berkeley, CA., 319-328.
- 1980, *The logic of evaluation*, Edgepress, Inverness, Ca.
- 1994, Evaluation as a discipline, *Studies in Educational Evaluation*, 20 (1), 147-166.
- STAKE, R. E., 1967, The countenance of educational evaluation, *Teacher College Record*, 68, 523-540.
- 1975, *Evaluating the arts in education: A responsive approach*, Merrill, Ohio.
- 1975a, Program evaluation: particularly responsive evaluation, *Occasional Paper*, 5, University of Western Michigan.
- 1998, A When policy is merely promotion, by what ethic lives an evaluator, *Studies in Educational Evaluation*, 24 (2), 203-212.
- STUFFLEBEAM, D. L., 1985, *Conducting Educational Needs Assessment*, Kluwer Nijhoff, Boston.
- 1999, Using professional standards to legally and ethically release evaluation findings, *Studies in Educational Evaluation*, 25 (4), 325-334.
- y otros, 1971, *Educational Evaluation and Decision-making*, F. E. Peacock Publishing, Itasca, Illinois.
- 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 Publications, Beverly Hills, Ca., 497-517.
- WEISS, C. H. 1970, The politization of evaluation research, *Journal of Social Issues*, 26 (4), 57-68.
- 1972, *Evaluation research: Methods of assessing program effectiveness*, Prentice Hall, Englewood Cliffs, N.J.
- 1973, The politics of impact measurement, *Policy Studies Journal*, 1 (3), 179-183.
- 1975, Evaluation research in the political context, en Struening, E. L. y Guttentag, M. (Eds.), *Handbook of Evaluation Research*, 1, Sage Publication, Beverly Hills, Ca. 13-26.
- 1998, Have we learned anything new about the use of evaluation?, *American Journal of Evaluation*, 19 (1), 21-33.
- WOLF, R. L., 1975, Trial by jury: a new evaluation method, *Phi Delta Kappa*, 57, 185-187.
- WORTHEN, B. R., SANDERS, J. R. y FITZPATRICK, J. L., 1997, *Program evaluation. Alternative approaches and practical guidelines*, Longman, New York.
- WRIGHT, C. R., 1967, Evaluation research, en *International Encyclopedia of the Social Sciences*, MacMillan, New York.