

1. APROXIMACIÓN CONCEPTUAL

M^a Teresa Anguera Argilaga
Salvador Chacón Moscoso

1.1. Perspectiva histórica

Berk y Rossi (1990) afirman que la evaluación de programas deriva del sentido común de que las intervenciones que se evalúan deberían ofrecer beneficios demostrables. Se halla implícita la idea de que los programas sociales y sanitarios tienen unos objetivos definidos, y serán juzgados empíricamente en función de su éxito o fracaso.

La evaluación de programas, considerada en sentido lato, tiene una larga historia, que, conjuntando aportaciones de diversos autores (Freeman y Solomon, 1981; Stufflebeam y Shinkfield, 1987; Shadish, Cook y Leviton, 1991; Fernández-Ballesteros, 1992a; Achúcarro y San Juan, 1996; Fernández del Valle, 1996b) diferenciamos en el largo período de antecedentes anteriores a 1930, la época de los grandes modelos desde 1930 a 1973, y la época actual de configuración como disciplina.

1.1.1. Antecedentes anteriores a 1930.

La evaluación, con o sin nombre, tiene una larga historia, y su existencia fue un hecho en distintas civilizaciones. No existen principios comunes, ni modelos que orienten el pensamiento o la realización empírica, y, por supuesto, falta esencialmente una sistematización.

A modo de ilustración, citamos algunos de estos antecedentes:

- En China, en el año 2000 a.C. se dirigieron investigaciones de los servicios civiles que no eran otra cosa que una evaluación de programas y de los individuos que los gestionaban (Bowman, 1989).
- En Grecia, en el siglo V a.C., Sócrates ya utilizó cuestionarios evaluativos.
- En la antigua Roma, las tasas que se cobraban por servicios públicos se modificaban según las fluctuaciones en los ingresos.
- En el siglo XVII se trataron de evaluar las causas de la desorganización social.
- A finales del siglo XVIII, la Armada británica promovía la bebida de zumos de cítricos en los largos viajes en la medida que prevenían enfermedades.
- En Inglaterra, en el siglo XIX se utilizaron comisiones reales para evaluar los servicios públicos.
- En 1845, en Estados Unidos, Mann dirigió una evaluación basada en tests de rendimiento para saber si las escuelas de Boston evaluaban bien a sus estudiantes.

- A finales del siglo XIX, en un estudio que está reconocido (Stufflebeam y Shinkfield, 1987) como la primera evaluación formal sobre un programa educativo realizado en América, se analizaron los conocimientos en ortografía de 33.000 alumnos.
- También a finales del siglo XIX se inició en Estados Unidos el movimiento para la acreditación de instituciones educativas y programas, por lo que debían evaluarse los servicios educativos.
- Ya en este siglo, en sus primeros años, por una parte se difundió profusamente el uso de los tests estandarizados.
- También a comienzos de siglo, la sentencia de prisión indeterminada se introdujo en los Estados Unidos de América como respuesta parcial a las elevadas tasas de reincidencia, por lo que los juicios se llevaban a cabo pretendiendo la efectividad de los programas en curso o prospectivos.

1.1.2. *Época de los grandes modelos entre 1930 y 1973.*

Durante bastantes años se sucedieron varios puntos de vista específicos sobre la evaluación, que fueron tejiendo la historia de la evaluación de programas, a la vez que se configuraban un gran número de modelos que hoy día nos invitan a una serena reflexión. Madaus, Scriven y Stufflebeam (1983) se apoyan en un trabajo clave de Smith y Tyler (1942) para el establecimiento de tres períodos en esta zona intermedia de la historia de la evaluación.

Dada la fuerte incidencia del ámbito sustantivo en el que se llevaba a cabo la evaluación, resulta obvia la existencia de una importante repercusión ejercida desde la evaluación educacional.

En los años de la década de los treinta e inicios de los cuarenta, la denominada *época tyleriana* se caracterizó por la conformación de un método centrado en la clara delimitación de unos objetivos, la confección de instrumentos estándar (tests diversos), y una definición de evaluación encaminada a la indagación sobre el cumplimiento de objetivos. Por primera vez, probablemente, no se necesita ya recurrir a comparaciones problemáticas entre grupos control y grupos experimentales, a la vez que adquieren consistencia los estudios relativos a diferencias individuales.

Durante el resto de años de la década de los cuarenta, así como durante los cincuenta, en un período caracterizado como *época de la inocencia*, o incluso *época de la irresponsabilidad social*, destaca una gran expansión en el acceso a recursos y un desorbitado consumo en determinadas zonas desarrolladas, aunque existiera segregación y profundos prejuicios raciales. Los servicios educativos se diversificaron, produciéndose un gran desarrollo en la educación, aunque no avances en la evaluación.

Hasta 1973 destaca un gran auge de evaluaciones de proyectos a gran escala, tanto en el ámbito educativo como social. A esta etapa, conocida como *época del realismo*, se le atribuyen grandes cambios guiados por el interés público y dependiendo

del dinero de los contribuyentes para su financiación, lo cual condujo a que la evaluación se convirtiera en una industria y en una profesión.

En el ámbito educativo ya no cabían evaluaciones al antojo del educador, pues los progresos de las metodologías evaluativas implicaban la priorización de criterios de utilidad y relevancia. En las evaluaciones realizadas en este período estaban presentes cuatro métodos evaluativos: a) El método de Tyler se utilizó para definir los objetivos de los nuevos *curricula* y para valorar el grado en que estos objetivos se lograban. b) Los nuevos tests estandarizados a nivel nacional sirvieron para reflejar mejor los objetivos y los contenidos de los nuevos currículos. c) El método del criterio profesional se usó para valorar las propuestas y verificar periódicamente los resultados por parte de asesores externos. d) Finalmente, muchos evaluadores se comprometieron a evaluar los resultados del desarrollo del currículo a través de pruebas concretas.

A pesar de este desarrollo, y de que se habían autoimpuesto la utilidad y la relevancia, éstas no se materializaban, y de ahí la valoración negativa de esta etapa por parte de Cronbach (1963, 1982). Su propuesta consistió en una reconceptualización no competitiva de la evaluación, sugiriendo que se configurase como un proceso consistente en recoger y formalizar información útil para quienes elaboran los currículos. Este esfuerzo obligaba a remodelar requisitos de la época tyleriana, y difícilmente se adaptarían a todo el ámbito nacional sus programas de tests estandarizados. Fue una época en que la inquietud existente por los malos resultados obtenidos actuó como revulsivo para el surgimiento de nuevas conceptualizaciones sobre la evaluación, las cuales reconocían la necesidad de evaluar las metas, examinar las inversiones efectuadas, y analizar el perfeccionamiento y la prestación de los servicios, a la vez que determinar los resultados que se desean obtener del programa.

Pero además, en los años transcurridos entre las dos Guerras Mundiales surgieron dos focos de evaluación de programas destinados a fines sociales con una orientación distinta entre sí:

Por una parte los de cuño caritativista, que pretendían resolver problemas carenciales de amplios colectivos poblacionales, y había que indagar cómo se empleaban los recursos respecto a la obtención de los objetivos que se pretendían, por lo que la evaluación y valoración de estos programas eran necesarias. En la sociedad norteamericana se tomó conciencia de las graves desigualdades existentes, iniciándose la puesta de marcha de importantes programas sociales que luchaban contra la pobreza, y que se sentían atraídos por las promesas de un Estado del bienestar. Como indica Fernández del Valle (1996a), los programas abarcaban prácticamente todos los sistemas de protección social (salud, educación, vivienda, etc.), pero la mayor novedad era de los programas de servicios sociales, como ayudas económicas, o protección familiar, o facilitación de residencias, etc.

Y otro núcleo era de carácter académico, en el cual investigadores sociales (psicólogos, sociólogos, demógrafos, etc.) contribuían al surgimiento de esta nueva disciplina de evaluación de programas, en la que se vertían los conocimientos de las

investigaciones aplicadas que realizaban. Había una preocupación intelectual que trataba de averiguar qué tipo de programas realizaban mejor su cometido o alcanzaban con mayor éxito sus objetivos. Destacan varios autores que en este período intermedio impulsaron decididamente la evaluación de programas, enfatizando el rigor científico necesario para la resolución de los problemas sociales y también sanitarios.

En conjunto, desde principios de los años sesenta, y en los Estados con importante desarrollo económico, como Estados Unidos, se pusieron en marcha un gran número de programas de intervención en el ámbito social (Bell, 1983), en su mayoría “a gran escala”, y desde una óptica caracterizada por preocupaciones metodológicas, dirigidos a mejorar la calidad de vida de sus ciudadanos. En su mayoría se iniciaron con muchas esperanzas, un gran despliegue de recursos humanos, y una intensa dotación económica (Shadish, Cook y Leviton, 1991).

Entre los diversos autores altamente relevantes de esta época por sus propuestas conceptuales y sus aportaciones metodológicas destacan especialmente:

- Suchman (1967), que realizó sus aportaciones en el ámbito de los programas de salud mental. Era un investigador que creía firmemente en la necesidad de basar principalmente las conclusiones en evidencias científicas, y que consideraba que la evaluación debía disponer de la lógica del método científico. En su importante obra dejaba constancia del necesario uso de todas las técnicas de investigación que fuesen útiles por parte de un evaluador. Distinguía entre evaluación e investigación evaluativa, y mientras consideraba a la primera como un proceso de emisión de juicios de valor, entendía que la segunda consistía en los procedimientos de recopilación y análisis de datos que aumentan la posibilidad de demostrar, más que de asegurar, el valor de alguna actividad social. La consideraba, por tanto, como investigación aplicada que debe permitir determinar hasta qué punto un programa específico ha conseguido el resultado deseado.
- Scriven (1967, 1972), defensor de la idea de que el evaluador, provisto de técnicas para obtener una información pertinente y exacta, y sin perder nunca de vista los principios éticos, debe ayudar a los profesionales a proporcionar productos y servicios de alta calidad y gran utilidad para los usuarios (“consumidores” en su terminología). Su espíritu crítico, desde el modelo de evaluación de meta libre, le llevó a defender insistentemente que, más que cumplir las metas propuestas, el evaluador debe juzgar si contribuye a aumentar el bienestar de los usuarios.
- Campbell (Campbell, 1969, 1970; Campbell y Stanley, 1963), pionero en la búsqueda de relación entre la experimentación y las reformas sociales, establecía básicamente que las reformas y programas sociales tendrían éxito si se realizaban las correspondientes investigaciones experimentales. Ha sido considerado uno de los defensores de la posición cuantitativa en evaluación en los momentos de máxima confrontación con la cualitativa. Creía en la existencia de un evolucionismo científico mediante el cual el conocimiento se acumula

selectivamente por su grado de rigor, y que aplicado a los programas sociales significaba la necesidad de una continua selección mediante valoraciones científicas.

1.1.3. Época actual de configuración como disciplina.

Hasta mitad de la década de los ochenta sigue desarrollándose la evaluación de programas, destacando diversos autores relevantes por su propuesta de modelos, como son:

- Weiss (1975a), que sitúa la evaluación de programas en un contexto político e inmerso en una dinámica social, cuestiones que el evaluador debe conocer en sus tomas de decisión. Consideraba de gran importancia las implicaciones que los resultados tenían en el desarrollo de los programas.
- Wholey (1979, 1989), muy pragmático, entendía que se realizaba una gestión deficiente de los programas sociales, lo cual les hacía ineficaces y difíciles de valorar. Proponía una implicación directa del evaluador en el programa, de forma que lo mejore a la vez que lo evalúa.
- Stake (1978) proponía igualmente que el evaluador debe involucrarse en el programa para realizar mejoras inmediatas, y puso un especial énfasis en las posiciones idiográficas y las metodologías flexibles.
- Cronbach (1980, 1982), preocupado por la necesidad de realizar una síntesis entre las distintas posiciones conceptuales sobre la evaluación, planteaba la importancia de identificar los elementos efectivos de los programas (utilizando el acrónimo 'utos', -unidades, tratamiento, resultados, contexto-, que se verá con más detalle en el capítulo tercero), con el fin de establecer generalizaciones que sean útiles.
- Rossi y Freeman (1989), finalmente, contemplaban una evaluación a medida, guiada por la teoría que se sustente, de forma contextualizada, abordando de igual forma el diseño del programa como su implementación e impacto, e introduciendo el análisis económico mediante la evaluación de costes en relación con la efectividad.

Se camina ya hacia un enfoque cada vez más integrador, signo inequívoco de mayor madurez y de que se adquiere la carta de naturaleza como disciplina. La evaluación de programas se consolida como tal cuerpo de conocimiento, adquiriendo flexibilidad para su aplicación a una amplísima cobertura de programas de intervención, capacitándose para incorporar todos los recursos metodológicos, y adquiriendo consistencia para prepararse como uno de los ámbitos punteros de la Psicología del 2000 (Wertheimer, Barclay, Cook, Kiesler, Koch, Riegel, Rorer, Senders, Smith y Sperling, 1978). En la actualidad, la consideración de la evaluación como profesión se

ha abierto camino, y se cuenta ya con interesantes reflexiones acerca de las alternativas que plantea (Modarresi, Newman & Abolafia, 2001).

El interés primordial se dirige a programas orientados a fines sociales de carácter diverso, como salud pública, empleo, o alimentación, incorporándose como temáticas nucleares las de prevención de la delincuencia y rehabilitación de sujetos encarcelados (Anguera y Redondo, 1991), y también programas educativos de variada casuística (Medina y Villar, 1995), programas de índole sanitaria (San Juan, García, Sota y Guimón, 1996) y hospitalaria (Echebarría y San Juan, 1996), así como de carácter social (Pérez Serrano, 1993) o comunitario (Sánchez Vidal, 1991). Asimismo, empiezan a subvencionarse investigaciones relativa a evaluación de programas o a encargarse proyectos de esta índole por parte de instituciones (Anguera y Blanco, 1988-91; Anguera, 1989; Anguera, Blanco, Gil, Ibern, Losada, Quera, Redondo y Roca, 1989-90; Chacón, Pérez-Gil, Holgado, Lara, López y Sanduvete 1998-2003; Chacón, Carmona, Pérez-Gil, Holgado y Losada, 2000-03; Anguera, Blanco, Losada, Sánchez Algarra, Hernández Mendo, Chacón, Ardá, Camerino y Castellano, 2001-04; Fernández-Ballesteros, 1984, 1986; Fernández-Ballesteros, Izal, Hernández, Montorio y Llorente, 1989).

La literatura científica, como caja de resonancia del avance en la cuestión, nos ofrece unas primeras aportaciones en los años setenta, principalmente provinientes de Norteamérica, y ya desde la década de los ochenta en España (como muestra, citamos a Fernández-Ballesteros, 1984, 1986, 1987, 1988, 1992a, 1992b, 1995; Alvira, 1985; Fernández-Ballesteros, Díaz-Veiga, Izal y González, 1987; Anguera, 1989, 1990; Ato, Quiñones, Romero y Rabadán, 1989; Caride, 1989; Fernández-Ballesteros, Izal, Hernández, Montorio y Llorente, 1989; Bartolomé, 1990; De la Orden, 1990; Mateo, 1990; Tejedor, 1990; Blanco, 1992; Carballo, 1992; García Ramos, 1992; Hernández y Rubio, 1992; Municio, 1992; Ramírez, 1992; Torroba, 1992; Chacón y López, 1993, 1995, 1998; Redondo, 1993; Reboloso, 1995; Casas, 1996; Fernández del Valle, 1996b; Medina, 1996; Anguera y Chacón, 1999; Chacón, Anguera, Pérez-Gil & Holgado, 2002). La eclosión es impresionante, y hasta se produce un cierto desgaste semántico del término por un uso que podría considerarse abusivo.

En el ámbito universitario está adquiriendo cuerpo progresivamente, y ya son varios los planes de estudios -al menos en las titulaciones de Psicología y Pedagogía- que lo contemplan, e incluso se propuso por primera vez en estudios de tercer ciclo¹ el año 1987. Institucionalmente, tanto si los recursos provienen del sector público como del privado, se va normalizando la evaluación (programas de atención a drogodependientes, de vacunación, de atención a neonatos, de niños institucionalizados, de calidad en la enseñanza, etc.), lo que conlleva la necesidad de una adecuada preparación en los profesionales. Pero, ¿se ha logrado esta formación?

¹ Programa de Doctorado *Evaluación en Ciencias del Comportamiento: Metodología y Aplicaciones*. Dpto. Metodología de las Ciencias del Comportamiento. Facultad de Psicología. Universidad de Barcelona.

La última etapa en evaluación de programas destaca por la profesionalización, lograda en parte, y tendente a alcanzar un desarrollo pleno en cuanto tal y la calidad técnica del programa (Greene, 1990). No en vano, Conner, Altmann y Jackson (1984) afirmaban hace poco más de una década que la evaluación se hallaba en la edad de la adolescencia.

La evaluación de programas indudablemente es compleja no sólo a nivel de contenido, sino procedimental, y también de estrategia (Linder & Peters, 1990), y su dificultad se halla condicionada por diversos factores, entre los que destacamos la naturaleza del programa de intervención, la existencia de instrumentos adecuados, los avances tecnológicos y los recursos disponibles. La formación metodológica, por otra parte, es más árdua, ya que debe implicar la posesión de un bagaje amplio de conocimientos para que el evaluador pueda llevar a cabo de forma certera las numerosas tomas de decisión que implica el proceso; por supuesto, no existe una pauta uniforme para el evaluador. Y, como telón de fondo, siempre permanece la pregunta relativa a la conexión -¿o relación causal?- entre las acciones que implica el programa y los resultados obtenidos (Toulemonde, 1995).

1.2. Posiciones actuales

Si preguntamos qué es evaluar, con toda seguridad obtendremos una gran diversidad de respuestas, y tanto más heterogéneas cuanto mayor sea la pluralidad profesional o procedencia formativa de quienes respondieran. Y entre los evaluadores de programas sociales y sanitarios es probable que también hayan discrepancias, e incluso dentro de cada uno de estos dos ámbitos. Pero como dicen Vega y Ventosa (1993), a pesar de la diversidad conceptual, todas tienen en común que evaluar es atribuir cierto valor al objeto de la evaluación. Tendremos que matizarlo, y para ello partimos de algunas definiciones posteriores a 1980, por ser a partir de entonces cuando la evaluación de programas empieza a ser una preocupación en España:

- Cronbach (1981) define la evaluación de programas como la provisión de información para la toma de decisiones sobre una intervención.
- La O.M.S. (1981) la caracteriza como un medio sistemático de aprender empíricamente y de analizar las lecciones aprendidas para el mejoramiento de las actividades en curso y para el fomento de una planificación más satisfactoria mediante una selección rigurosa entre las distintas posibilidades de acción futura. Ello supone un análisis crítico de los diferentes aspectos del establecimiento y la ejecución de un programa y de las actividades que lo constituyen, su pertinencia, su eficiencia y eficacia, su costo y su aceptabilidad para todas las partes interesadas.

- La O.N.U. (1984) la define como el proceso encaminado a determinar sistemática y objetivamente la pertinencia, eficiencia, eficacia e impacto de todas las actividades a la luz de sus objetivos. Se trata de un proceso organizativo para mejorar las actividades todavía en marcha y ayudar a la administración en la planificación, programación y toma de decisiones futuras.
- House (1984) vincula la evaluación a la metáfora subyacente en el evaluador, conectada a experiencias culturales y a otros ámbitos ajenos a la evaluación. Los programas sociales se pueden cuantificar y medir mediante su conversión metafórica en objetos, los cuales se pueden describir, pero en esta operación caben muchas diferencias según su tamaño, cobertura, dimensiones, ..., y aquí es donde juega su papel la naturaleza metafórica de la evaluación.
- Posavac y Carey (1985) consideran que la evaluación es una nueva ciencia social que pretende la mejora de los servicios humanos que se presten, y que merced a un conjunto de métodos, competencias y sensibilidades necesarias se determina si el servicio es necesario y cómo se dispensa, a la vez que se estudia si es suficiente.
- Alvira (1985) considera que es simplemente la acumulación de información sobre una intervención -programa-, sobre su funcionamiento y sobre sus efectos y consecuencias. Lo matiza posteriormente (Alvira, 1991) diciendo que evaluar es emitir juicios de valor, adjudicar valor o mérito a un programa/intervención, basándose en información empírica recogida sistemática y rigurosamente.
- Stufflebean y Shinkfield (1987) afirman que evaluar es el enjuiciamiento sistemático de la valía o el mérito de un objeto. O, de forma más elaborada, el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los fenómenos implicados.
- Cabrera (1987) retoma la tradición de Cronbach y Scriven, y define la evaluación como el proceso de recogida y análisis de información relevante en que apoyar un juicio de valor sobre la entidad evaluada que sirva de base para una eventual toma de decisiones.
- Pineault y Daveluy (1987) afirman que la evaluación es un proceso que consiste en determinar y en aplicar criterios y normas con el fin de realizar un juicio sobre los diferentes componentes del programa, tanto en el estadio de su concepción, como de su ejecución, así como sobre las etapas del proceso de planificación que son previas a la programación.

- Caride (1989) la entiende como el análisis e interpretación valorativa de los resultados y logros del programa y sus correspondientes procesos. Consiste en un proceso de reflexión y valoración crítica, continua y sistemática, referida a momentos y factores que intervienen en el desarrollo de un programa, al objeto de determinar cuáles pueden ser, están siendo o han sido sus efectos, resultados o logros.
- En su interesante discusión sobre evaluación, Alkin, Patton, Weiss, Conner, House, Kean, King, Klein, Law y McLaughlin (1990) perfilan el concepto, que se puede entender como la actividad de recoger información sistemáticamente, analizarla, y conocer consecuentemente en qué medida se ha producido un cambio (posición de Alkin), así como también la asignación de un determinado valor según una serie de criterios o estándares explícitos o implícitos (posición de House). A su vez, Kean considera que es el proceso de delimitar, obtener y elaborar información útil para juzgar entre decisiones alternativas. Por su parte, Weiss considera que es un tipo de estrategia de investigación que ofrece una evidencia sistemática gracias a observaciones, encuestas, entrevistas no estructuradas, tests, cuestionarios, inventarios, material documental, y reanálisis de material de archivo, todo lo cual ayudará a tomar decisiones con imparcialidad, pero no a generalizar.
- Fernández-Ballesteros (1992a) afirma que consiste en una sistemática investigación del mérito, valor o éxito de concretos tratamientos, intervenciones o programas llevados a cabo con el fin de producir cambios comportamentales en el contexto social. Expresado de forma parecida (Fernández-Ballesteros, 1995), evaluación de programas es la sistemática investigación a través de métodos científicos de los efectos, resultados y objetivos de un programa con el fin de tomar decisiones sobre él.
- Holfstee (1992) plantea la evaluación de programas desde la perspectiva del usuario de servicios públicos, en los cuales los recursos son aportados por el conjunto de los ciudadanos. Por este motivo plantea la definición desde la complementación coste-beneficio y coste-efectividad, por una parte, y, por otra, desde el esfuerzo por lograr la calidad en los servicios.
- Cohen y Franco (1992), después de discutir otras definiciones, la contextualizan en su obra considerándola como una actividad que tiene por objeto maximizar la eficacia de los programas en la obtención de sus fines y la eficiencia en la asignación de recursos para la consecución de los mismos. Matizan que evaluar es fijar el valor de una cosa; para llevarlo a cabo se requiere efectuar un procedimiento mediante el cual se compara aquello a evaluar respecto de un criterio o patrón determinado.

- Aguilar y Ander-Egg (1992) consideran que “la evaluación es una forma de investigación social aplicada, sistemática, planificada y dirigida; encaminada a identificar, obtener y proporcionar de manera válida y fiable, datos e información suficiente y relevante, en que apoyar un juicio acerca del mérito y el valor de los diferentes componentes de un programa (tanto en la fase de diagnóstico, programación o ejecución), o de un conjunto de actividades específicas que se realizan, han realizado o realizarán, con el propósito de producir efectos o resultados concretos; comprobando la extensión y el grado en que dichos logros se han dado, de forma tal, que sirva de base o guía para una toma de decisiones racional e inteligente entre cursos de acción, o para solucionar problemas y promover el conocimiento y la comprensión de los factores asociados al éxito o al fracaso de sus resultados” (p. 18).
- Vedung (1993) la define como un mecanismo *ex-post* para la exposición y valoración de la implementación, resultados y logros de las políticas y programas públicos con el propósito de generar posteriores decisiones.
- Fernández del Valle (1996b), además de considerar que la evaluación de programas es una exigencia metodológica para todos los programas de intervención, entiende que cubre una triple exigencia: En primer lugar es una exigencia profesional, ya que es responsabilidad de todos los que intervienen en los programas el conocer y estudiar sus propios procesos y resultados de intervención con el fin de mejorar su ejercicio; en segundo lugar, es una exigencia científica en cuanto que la intervención psicosocial es un campo en construcción que demanda un desarrollo teórico y unos fundamentos que necesariamente pasan por analizar rigurosamente las prácticas desarrolladas, contrastando y comprobando las nuevas aportaciones técnicas y teóricas; y finalmente, la evaluación de programas es una exigencia ética ya que permite el control de los efectos causados con una determinada intervención, y por tanto, garantiza la adecuación a ciertos principios prescriptivos como la igualdad de oportunidades, el respeto a las minorías, etc.

La reflexión sobre estas aportaciones particulares nos permite afirmar que nos hallamos en un momento de síntesis, lo cual valoramos positivamente por cuanto que aún a consideraciones diversas. Se complementan las preocupaciones por situar las valoraciones en el contexto político y administrativo con la recuperación del interés inicial por la rigurosidad del método y sus garantías científicas (Fernández del Valle, 1996b), a la vez que se ha conformado una disciplina altamente estructurada que

permite conducir de forma concatenada las diversas fases del proceso (Herman, Morris & Fitz-Gibbon, 1987; Brown & Kiernan, 2001).

Una cuestión no exenta de polémica y que consideramos muy relevante es la diferencia entre el uso de los términos *evaluación* y *valoración*. Remitimos al lector a los trabajos de Fernández-Ballesteros (1985, 1995).

En el primero de ellos, se dice que “en lengua inglesa, a pesar de que los vocablos *assessment* y *evaluation* son sinónimos, presentan, en la literatura especializada, usos diferentes. Tratando de ser fiel a esta diferenciación semántica y traduciendo *evaluación* por *assessment* y *valoración* por *evaluation* es por lo que en distintos trabajos me he referido (...) a dos momentos distintos: Uno inicial de evaluación propiamente dicha, en el que se trata de proceder al análisis de los comportamientos problema así como de las variables ambientales y/o internas que los controlan o mantienen con el fin de proceder a la formulación de hipótesis funcionales que permitan, a su vez, seleccionar un determinado tratamiento. A partir de este momento, se iniciaría un proceso valorativo del propio tratamiento aplicado que va a dar como resultado un juicio sobre el valor, mérito o eficacia de ese concreto tratamiento y que, por ende, va a permitirnos verificar, o no, las hipótesis funcionales formuladas” (Fernández-Ballesteros, 1985, p. 8).

En Fernández-Ballesteros (1995, p. 22) reaparece esta cuestión terminológica, matizándose que “evaluación es un término menos cargado semánticamente que valoración. Ello tal vez justifica (si no explica) el hecho de que la literatura en español de la disciplina que nos ocupa aparezca bajo la rúbrica de *evaluación* y no de *valoración* como sería más preciso ateniéndonos a los usos matrices en inglés”.

Aquí seguimos esta tónica terminológica, y mantenemos el término *evaluación* para referirnos al estudio científico del proceso. Estamos de acuerdo con Fernández del Valle (1996a, p. 7) cuando afirma que “la evaluación de programas (debería decirse valoración de programas para ser más rigurosos) supone en la actualidad una exigencia metodológica para todos los programas de intervención”. No obstante, si finalizado el proceso es necesaria una referencia a la emisión de juicios de valor, utilizaremos el término *valoración*.

1.3. Modelos de evaluación

Todo proceso evaluativo ha de tener presente, en alguna medida, un modelo teórico en el que basarse, ya que toda evaluación se caracteriza por un continuo de toma de decisiones dónde se necesita disponer de un referente teórico a la hora de orientarla. En este sentido de acuerdo con lo presentado en los apartados y considerando el orden que plantea Cook (1988) se pueden observar tres fases generales en la evolución de las teorías de evaluación.

Scriven y Campbell en los años sesenta representarían la primera fase; en esa época se defendió la evaluación como un proceso de investigación rigurosamente científico, regido por principios de causa-efecto lineales, desde los que plantear intervenciones sociales eficientes con las que resolver los problemas sociales.

En una segunda fase, debido fundamentalmente al descontento con los resultados obtenidos mediante los planteamientos anteriores, se produjo la búsqueda de otras alternativas, pudiendo mencionarse entre otros autores las aportaciones de Weiss (1975b) que prestaba mayor atención a la modulación del contexto sociopolítico; por su parte Stake (1975) enfatizaba la importancia de los promotores locales de los programas; Wholey (1977) planteaba la necesidad de valorar la evaluabilidad del programa; Patton (1978) priorizaba el uso que se le da a los resultados del proceso evaluativo, mientras Guba (1987) enfatizaba el análisis de los distintos posibles intereses que confluyen en el programa mediante métodos cualitativos.

En último término, la etapa denominada de integración, donde se han intentado coordinar los distintos planteamientos existentes, podría estar representada por autores como Chen y Rossi (1983), con sus planteamientos evaluativos multipropósito (a partir de distintos modelos teóricos) y de teoría-dirigida (toda intervención ha de basarse en algún modelo teórico, explícito o no), al igual que Cronbach (1982), que enfatizaba la validez externa del proceso evaluativo, pudiendo incluirse en este grupo a Conner (1984) con sus planteamientos metodológicos integradores, principalmente orientados a la implementación del programa de evaluación.

Con la presentación de estos estadios en la evolución teórica de la evaluación de programas entendemos que se corre el riesgo de simplificar la complejidad de la realidad evaluativa, en tanto que lejos de tratarse de un proceso caracterizado por la sucesión de etapas completamente excluyentes, nos referimos a un continuo interdependiente en el que los últimos teóricos asumen, eliminan o modifican parte de los trabajos anteriores.

Esta situación provoca que algunos de los autores citados se resistan a ser clasificados sólo dentro de una única etapa -de hecho los planteamientos de Wholey y Weiss asumen en ciertos aspectos la aproximación de Campbell-; de igual forma, Cronbach en 1963 ya adelantaba sus planteamientos integradores en la primera fase dominada por los modelos objetivistas, donde se enfatizaba principalmente la consecución de objetivos.

Esta misma circunstancia se pone de manifiesto incluso en la propia trayectoria de algunos de los autores citados como por ejemplo Stake, Weiss o Cronbach, donde se observa cómo dentro de sus propios planteamientos teóricos han ido evolucionando desde posturas más rígidas, características de los modelos "experimentalistas", a perspectivas más flexibles y eclécticas.

De igual forma, hablar de una última etapa de integración sería cierto si sólo hubiese un único planteamiento; en cambio, los teóricos de este estadio tampoco

coinciden plenamente, estando hoy día abierta la polémica acerca de qué aspectos relevantes considerar en el proceso de evaluación.

A pesar del riesgo de la "sobresimplificación", defendemos que los tres estadios reflejan una tendencia general que entendemos puede ser útil para anclar el análisis metodológico del proceso evaluativo, que contemplamos desde la vertiente investigadora (Crane, 1988; Chacón, Cook, Pérez-Gil, Anguera & Shadish, 2001; Chacón & Shadish, 2001) y aplicada (Anguera y Redondo, 1991; Chacón, Pérez-Gil, Holgado y Lara, 2001; Chacón, Pérez-Gil & Holgado, 2001).

1.4. Tipos de evaluación

La amplia diversidad de programas de intervención a evaluar, así como los presupuestos de los cuales se parte, contribuyen a modular diversas modalidades evaluadoras.

La taxonomía en evaluación de programas se puede enfocar con criterios múltiples. En gran medida estamos de acuerdo con las de Aguilar y Ander-Egg (1992) y Fernández-Ballesteros (1995), a la vez que consideramos interesantes aspectos incorporados por López Blasco (1992), especialmente referidos a la cobertura del programa, por Muñoz Tortosa (1992), teniendo en cuenta las fases del seguimiento, y por Cohen y Franco (1992) y Reboloso (1994), atendiendo a diferentes criterios complementarios que presentamos a continuación.

1. Según las funciones que cumple: Evaluación formativa y sumativa.

Se trata de la distinción que de forma más unánime recogen los diversos autores (Veney y Kaluzny, 1984; Aguilar y Ander-Egg, 1992; Mohr, 1992; Cohen y Franco, 1992; Reboloso, 1994; Fernández-Ballesteros, 1995; Shiratuddin & Landoni, 2002) a partir de la propuesta de Scriven (1967, 1991).

La *evaluación sumativa* hace referencia a los resultados o efectos de un programa, una vez finalizado, determinándose así hasta qué punto se cumplen los objetivos o se producen los efectos previstos. Determina el valor de un programa una vez que ha sido desarrollado, investiga los efectos y los compara con las necesidades de los usuarios. Veney y Kaluzny (1984) consideran esta evaluación situada en el *modelo lineal* en la gestión de un programa, que se planifica, implementa, y, finalmente, evalúa, como fase última que no admite retorno. También recibe el nombre de *evaluación de producto*, o *de resultados*, o *de impacto* (Mohr, 1992), aunque se considera que la información de que se dispone es insuficiente para poder cuantificar el verdadero impacto del programa, del que raramente se señalan subobjetivos. Fundamentalmente comprende actividades de análisis de datos.

Son tareas básicas de este tipo de evaluación (Reboloso, 1994):

- a) Recopilar descripciones y juicios sobre los resultados.
- b) Relacionarlos con los objetivos y la información proporcionada por el contexto, por la entrada de datos y por el proceso.
- c) Analizar e interpretar su valor y mérito.
- d) Proporcionar información objetiva y clara sobre los resultados deseados y no deseados.

Los datos de una evaluación sumativa suministran información útil para la toma de decisiones sobre la continuación, justificación, expansión o reducción, contabilidad, modificación y finalización de un programa. Pero habitualmente suscita importantes problemas metodológicos, como es el hallar indicadores adecuados del impacto, o poder atribuir los supuestos efectos del programa a las variables independientes fijadas.

La *evaluación formativa* hace referencia al seguimiento que se realiza durante el proceso de implementación de un programa (Fitzpatrick, 1988), proporcionando información acerca de cómo se desarrolla dicho proceso, y sirve básicamente para ayudar a la puesta en marcha de un programa o para mejorar lo que se está realizando o su funcionamiento (Dehar, Casswell & Duignam, 1993; Tessmer, 1994). Veney y Kaluzny (1984) consideran esta evaluación integrada en el modelo no lineal de gestión de un programa, dado que la planificación, la implementación y el control se interconectan, por lo que la evaluación tiene lugar durante todas las fases del proceso. En cualquier momento se evalúa la necesidad, cómo se actúa y la suficiencia de las acciones emprendidas, los problemas que surgen en el día-a-día, los tipos de datos que más conviene obtener, los recursos necesarios, etc. La información que se suministra es periódica, o incluso continua, aunque en ocasiones se admite por algunos autores (Mohr, 1992) que pueda ser de carácter adicional, aunque conservando su aportación periódica o continua.

Las actividades cubren un amplio margen de actuaciones:

- a) Análisis de estrategias de interacción entre las personas involucradas en el programa.
- b) Periodificación en la implementación.
- c) Agregación de datos a lo largo del seguimiento.
- d) Integración de informaciones heterogéneas
- e) Comprobación del cumplimiento del desarrollo y proceso del programa.

Autores como Cronbach (1981) piensan que los dos términos de *evaluación sumativa* y *evaluación formativa* no son adecuados para la discusiones de hoy en día, pues las evaluaciones se usan casi totalmente de un modo formativo, incluso en los estudios que se autolimitan a medir resultados.

Otros, en cambio, como Aguilar y Ander-Egg (1992), o Fernández-Ballesteros (1995) entienden que no hay que menospreciar ningún tipo, pues las dos son complementarias (Chambers, 1994), y cada una de ellas puede utilizarse válidamente, en función del mayor o menor grado de elaboración del programa, así como del contexto y la situación concreta.

Tradicionalmente se siguió en España la evaluación sumativa con mayor profusión, lo cual concuerda con la praxis de fijación de plazos en la implementación de programas que dependen de alguna Administración Pública (estatal, autonómica o local), pero paulatinamente aumenta el interés por la evaluación formativa, dadas las ventajas que aporta desde su abordaje metodológico (corregir un inadecuado uso de determinado instrumento, introducir datos complementarios en el registro, etc.), así como la posibilidad de ajustar el PERT (*Project Evaluation Review Technique*) a lo largo de la implementación del programa.

2. Según el momento en que se evalúa: Evaluación “*ex-ante*”, durante y “*ex-post*”.

La distinción basada en este criterio contiene las tres posibilidades:

La *evaluación “ex-ante”* (llamada también *antes*, o *inicial*, o *pre-decisión*) se realiza antes de aprobar la realización de un programa, o antes de emprender su implementación, y, por tanto, en la fase de diseño del programa. Consiste en evaluar el programa en sí mismo, mediante la estimación crítica de su pertinencia, viabilidad y eficacia potencial, con el fin de proporcionar criterios racionales sobre la conveniencia o no de llevarlo a cabo. Según Pineault y Daveluy (1987), trata de comparar la eficacia o la rentabilidad económica de las diferentes acciones para alcanzar los objetivos deseados. Aquí se ubican los programas de carácter preventivo, tan útiles en ocasiones en que ya se ha efectuado una evaluación de necesidades futuras, o en que es pública la existencia de una carencia o peligro (Leviton & Valdiserri, 1990; Berk & Schulman, 1995).

La *evaluación durante* (denominada también *de gestión*, *continua*, *concurrente*, *concomitante*) es la que se realiza durante la fase de ejecución o implementación, suministrando información sobre la marcha del programa con una ponderación de los resultados. Su objetivo básico es evaluar los cambios situacionales, es decir, conocer en qué medida se está cumpliendo y realizando el programa o prestando un servicio de acuerdo a la propuesta inicial.

La *evaluación “ex-post”* (que debe diferenciarse de la evaluación de fin de proyecto) se lleva a cabo (Aguilar y Ander-Egg, 1992) cuando el programa ha alcanzado su pleno desarrollo, que puede ser meses o incluso años después de finalizada la ejecución. Se trataría de una modalidad de evaluación de resultados, y que habitualmente se realiza para evaluar cuánto y cómo ha cambiado la situación inicial, o cuándo se ha logrado la situación objetivo, según lo que se tome como punto de referencia). En numerosas ocasiones se utilizan datos retrospectivos (Holland & Rubin, 1998) o de archivo (Lucky, Broughton & Sorensen, 1983).

3. Según la procedencia de los evaluadores: Externa, interna, mixta, autoevaluación, evaluación participativa.

En sus rasgos esenciales existe igualmente consenso entre los distintos autores.

La *evaluación externa*, o *desde fuera*, es la realizada por personas ajenas a la organización del programa, en tanto que no pertenecen ni están vinculadas (directa o indirectamente) a ella. Se incrementa la objetividad, y se plantea la polémica sobre si tiene más relevancia un arsenal metodológico más sofisticado -no en vano se maximiza la utilización de tecnología dura- o el conocimiento sustantivo al que se refiere el programa. Los críticos de la evaluación externa se preguntan si realmente existe un arsenal metodológico único, a la vez que plantean dudas a la objetividad del evaluador, aduciendo que no siempre está garantizada, y alegan el mayor coste que supone, así como su carácter 'controlador' (Thompson, 1981); por el contrario, los defensores argumentan la mayor credibilidad social que comporta, además de aumentar la objetividad. De ordinario se acude a la participación de expertos contratados por la institución que quiere evaluar el programa que desarrolla.

La *evaluación interna*, o *desde dentro*, supone una implicación del evaluador al pertenecer a la organización gestora del programa, aunque no es directamente responsable de la implementación del programa. Se aduce, como aspecto positivo, que se reducen las fricciones propias de la evaluación externa, puesto que no sería percibida como control, sino como proceso de reflexión sobre lo realizado. Se supone que el evaluador conoce lo sustantivo de la disciplina, y que ejerce una mayor influencia sobre el programa, además de que los costes de la evaluación se reducen. Pero a nadie escapa que el ser a la vez juez y parte reduce las garantías de objetividad.

La *evaluación mixta* busca combinar los dos tipos anteriores, haciendo que evaluadores externos realicen su trabajo en estrecho contacto y con la participación de miembros del programa a evaluar. De esta manera se pretende superar las dificultades, a la vez que se preservan las ventajas de ambas. Probablemente sea la más conveniente en la mayoría de los casos, al lograr equilibrar los factores desfavorables y reforzar los favorables. Además, permite evitar la oposición frontal entre las perspectivas externa e interna de la evaluación, matizándose los niveles de implicación del evaluador.

La *autoevaluación*, mencionada por muy pocos autores (Aguilar y Ander-Egg, 1992), tiene lugar cuando las actividades evaluativas están a cargo de los directamente implicados en la ejecución del programa, y que valoran y enjuician sus propias actividades para determinar si están cumpliendo las metas propuestas (Folkman & Rai, 1997). Cada profesional puede evaluar su competencia en el desempeño de su profesión (Lysaght & Altschuld, 2000).

La *evaluación participativa* surgió para minimizar la distancia que existe entre el evaluador y los beneficiarios (Aguilar y Ander-Egg, 1992), y se ha incrementado notablemente su filosofía en los últimos años, aunque también tiene sus detractores (Gregory, 2000). Se utiliza fundamentalmente en programas de pequeña envergadura que buscan asentar los cambios propugnados mediante la creación de condiciones idóneas para que se genere una respuesta endógena del grupo (Robinson y Whitfield,

1987; Whitmore, 1991; Mathie & Greene, 1997; Levin, 1999). Prevé el concurso de la comunidad (población diana del programa) en la programación de la intervención, así como en la implementación y diseño de la evaluación.

4. Según los aspectos del programa que son objeto de evaluación.

Es un criterio con importantes repercusiones metodológicas, pero también con diferencias notables entre los distintos autores, ya que los subcriterios no coinciden (Aguilar y Ander-Egg, 1992):

- Rutman (1977) y Rossi y Freeman (1989) diferencian la evaluación del diseño y conceptualización del programa, la evaluación o seguimiento de la instrumentación y desarrollo del programa, y la evaluación o valoración de la efectividad y eficiencia del programa.
- Donabedian (1980) y Pineault y Daveluy (1987) distinguen entre la evaluación de la estructura, del proceso y de los efectos.
- Luque y Casquete (1989) consideran que las evaluaciones posibles son las de necesidades, de procesos, de resultados, de eficacia y de costos.
- Caride (s/f) distingue la evaluación de contexto, de necesidades, del diseño, del proceso y del producto.
- Alvira (1991) establece la evaluación de necesidades, del diseño / conceptualización del programa, de la evaluabilidad, de la implementación, de la cobertura, de la monitorización y seguimiento del programa, de resultados, de impacto y económica.
- Aguilar y Ander-Egg (1992) distinguen la evaluación del diseño y conceptualización del programa, la de la instrumentación y seguimiento del programa (evaluación del proceso), y la de la eficacia y eficiencia del programa (evaluación de resultados).

5. Según los fines de la evaluación: Evaluación proactiva y evaluación retroactiva

Fernández-Ballesteros (1995) adapta una tipología establecida por Stufflebeam y Shinkfield (1987) según los propósitos de la evaluación.

La *evaluación proactiva* implica basarse en la toma de decisiones del programa, normalmente con el fin de mejorar dicho programa.

La *evaluación retroactiva* parte de la información recogida durante el tiempo de funcionamiento para dar cuenta de ello como indicador de resultados que es necesario para la evaluación económica, así como para la nueva determinación del presupuesto para la puesta en marcha del programa.

6. *Evaluación formal y sustantiva*

De forma divergente con el resto, esta clasificación atiende a las áreas que son evaluadas, y tuvo su momento de mayor aceptación en la década de los ochenta. La comparten Espinoza (1986) y Aguilar y Ander-Egg (1992).

La *evaluación formal* atiende a los componentes internos del programa, y se juzga la pertinencia potencial del programa o coherencia interna entre sus distintos componentes. Las siete áreas de evaluación que incluyen son: Estudio y diagnóstico, diseño o fase de programación, proceso y desarrollo del programa (comparar lo que se está realizando con lo planificado), recursos (cantidad, calidad y disponibilidad de recursos para implementar el programa), perfil estructural y funcional de la institución desde la cual se realiza el programa, de los procedimientos utilizados, y de la coordinación del programa con otros.

La *evaluación sustantiva* trata de lograr la pertinencia de los logros del programa en relación con los problemas o necesidades que afectan a la población destinataria y que dieron lugar a la intervención. Se juzga la pertinencia real del programa. Las tres áreas de coherencia externa son: evaluación de los resultados efectivos, opinión de los usuarios o destinatarios del programa, y evaluación de los efectos indirectos y de los no esperados.

1.5. Componentes

Consideramos componentes de un programa a evaluar aquellos criterios de calidad que, tanto si se contemplan de forma individualizada como agrupada, le aportan idiosincrasia y permiten caracterizarlo para su posterior evaluación, a la vez que se convierten en su garantía de control.

Presentamos a continuación los que juzgamos más relevantes:

- *Aceptabilidad para las partes interesadas.*
Medida en que cada uno de los colectivos o personas implicadas (responsable de la institución, profesional, usuario, allegado, etc.) desempeña el papel que tiene asignado, sin interferencias entre ellos. Un apartado de este componente lo constituye el acceso a los beneficios del programa (Pandiani, Banks, Bramley, Pomeroy & Simon, 2002), así como la admisión por parte de los responsables del programa (Garner, Godley & Funk, 2002).
- *Adecuación*
La adecuación trata de asegurar que una evaluación revelará y expresará información técnicamente adecuada sobre aquellas características que determinan valor o mérito del programa (Fernández-Ballesteros, 1995; Joint Committee, 1994).
- *Calidad o bondad*
La naturaleza y planteamiento de cada una de las acciones del programa se desarrollan de forma adecuada según los criterios técnicos y de expertos.
- *Cobertura*
Caracterización del colectivo de usuarios y de potenciales usuarios en función de las necesidades o carencias detectadas previamente (Alvira, 1991; Cohen y Franco, 1992).
- *Coherencia*
La coherencia y congruencia internas es un análisis que atiende fundamentalmente a los aspectos formales que evalúan la organicidad/unidad del programa (medios/recursos y objetivos, objetivos generales, objetivos específicos y metas, diagnóstico y proposiciones), así como al hecho de que sea sostenible como tal (Van Vlaenderen, 2001).
- *Continuidad*
Establecer si se mantiene la formulación inicial o se redefinen los objetivos y metas propuestos y las acciones programadas.
- *Efectividad*
Medida en que se generan efectos de algún tipo (positivos, negativos, neutros, ...) como consecuencia de la implementación de un programa (Artells, 1989; Fernández-Ballesteros, 1992a, 1995; OMS, 1981).
- *Eficacia*
Cumplimiento de los objetivos especificados, tanto cuantitativamente (todos ellos) como cualitativamente (en la medida en que se había prefijado) (Cochrane, 1985; Fernández-Ballesteros, 1992a, 1995; OMS, 1981; Flay, 1986; Pineault y Daveluy, 1987).

- *Eficiencia*
Logro del mismo grado de eficacia con menos recursos, o con un mayor abaratamiento de éstos (Artells, 1989; Cochrane, 1985; Cohen y Franco, 1992; Fernández-Ballesteros, 1992a, 1995; OMS, 1981).
- *Equidad*
Medida en que la planificación (Krumholz, 1982) y la implementación de un programa se efectúa con iguales estándares para todos los individuos que integran el colectivo de usuarios / potenciales usuarios, y que, consecuentemente, gozan de igualdad de oportunidades (Artells, 1989; Cohen y Franco, 1992). También puede aplicarse al análisis económico de programas (Ubel, DeKay, Baron & Asch, 1996).
- *Evaluabilidad*
Medida en que un programa puede ser evaluado desde los distintos criterios que interesen (Wholey, 1977, 1989; Smith, 1990; Alvira, 1991; Fernández-Ballesteros, 1988, 1992b, 1995; Fernández-Ballesteros y Hernández, 1995; Hernández y Rubio, 1992; Leviton, Collins, Laird & Kratt, 1998), incluso en programas preventivos (Gibbs, Napp, Jolly, Westover & Uhl, 2002).
- *Factibilidad*
Medida en que la implementación del programa es posible en función de las acciones previstas (Anguera, 1990; Fernández-Ballesteros, 1995; Joint Committee, 1994; Pineault y Daveluy, 1987).
- *Pertinencia*
Coherencia entre las soluciones que se proponen y la realidad que se pretende modificar (Fernández-Ballesteros, 1995; OMS, 1981).
- *Probidad*
Grado en que la evaluación se ha realizado de forma ética y legal, y con la debida atención al bienestar de los usuarios y demás implicados en la evaluación (Fernández-Ballesteros, 1995; Joint Committee, 1994).
- *Progreso*
Evidencia de la progresiva implementación del programa como consecuencia del cumplimiento de objetivos (Fernández-Ballesteros, 1995; OMS, 1981).
- *Suficiencia*
Previsión de que las acciones previstas en el programa son capaces de resolver la necesidad existente (Fernández-Ballesteros, 1995; OMS, 1981).

- *Usabilidad*
Medida en que se acude al programa preestablecido para resolver una determinada necesidad en un determinado individuo o colectivo (Keinonen, 1998). La utilización conceptual de un programa se puede modular en función del marco teórico (Johnson, 1998).
- *Utilidad*
Influencia directa y automática de los resultados en los programas en la posterior reflexión (Esteve, Godoy, Rodríguez-Naranjo y Fierro, 1993; Fernández-Ballesteros, 1995; Joint Committee, 1994; Qvortrup, 1992; Vedung, 1993, 1995).
- *Viabilidad*
Disponibilidad de los mínimos recursos necesarios para que funcione el programa (Ibern y Anguera, 1990).

La consideración conjunta de estos componentes nos lleva a perfilar de forma complementaria, una serie de cuestiones, como las características que debe reunir una buena evaluación, cómo no debe ser, cómo incide en ella la política, y cuáles son las funciones sociales de la evaluación:

A. Cómo debe ser una buena evaluación

El Joint Committee on Standards for Educational Evaluation (1981) se refiere a unas condiciones básicas que ha de cumplir una buena evaluación. A continuación se presentan las más relevantes (Fernández-Ballesteros y Hernández, 1995):

- a) *Que sea útil*, es decir, que esté en relación y se corresponda con lo que necesitan las personas implicadas en el programa a evaluar (responsables, profesionales, participantes, usuarios de la evaluación o “clientes”).
- b) *Que sea factible o viable*, pudiéndose realizar en la práctica dentro de las posibilidades y la coyuntura que ofrezcan las circunstancias del momento.
- c) *Realismo*, ajustándose a los recursos disponibles.
- d) *Prudencia y moderación*, sin que resulte “peor el remedio que la enfermedad que se pretende curar”. Antes de aplicar una evaluación conviene calcular y prever si sus resultados van a traer más ventajas (mejora del programa, de las relaciones, del clima de trabajo, etc.) que inconvenientes (reacciones negativas, conflictos, frustraciones, ansiedad, desánimo, etc.).
- e) *Diplomacia*. La evaluación ha de saberse llevar, tanto en su diseño como en su administración, así como en la difusión de los resultados. No se trata de falsear o escamotear cuestiones críticas, pero sí el plantearlas con delicadeza, respeto, y sobre todo con una actitud positiva y constructiva.
- f) *Eficiencia*. La viabilidad de una evaluación también se mide en función de la relación existente entre los resultados pretendidos y los recursos o medios requeridos.
- g) *Que sea ética*, basada en la cooperación y protección de los derechos de las partes implicadas en la evaluación (institución, equipo de trabajo, participantes). Existe la obligación moral de devolver a la comunidad, de forma sistematizada (mediante el informe final) aquello que se ha extraído de ella (información), y sin ningún tipo de manipulación.
- h) *Que sea precisa o exacta*. Se requiere una buena pericia y preparación de los evaluadores para prever y controlar una gran cantidad de factores que intervienen. En muchos casos, la compleja realidad con la que se trabaja hace imposible el control o aislamiento de numerosas e impredecibles variables que influyen y contaminan los resultados; sin embargo, sí que se deberán utilizar procedimientos con los que poder minimizar o prever en lo posible los factores distorsionantes.

B. *Cómo no debe ser una evaluación.*

Existen una serie de vicios o errores usuales con los que nos encontramos frecuentemente, y que conviene desterrar de la práctica evaluativa. Vega y Ventosa (1993) los comentan en evaluaciones de animación sociocultural, pero aquí los planteamos referidos a la generalidad de los casos:

- a) *Pseudoevaluaciones*, cuando no se lleva a cabo una auténtica evaluación por no haberse planteado previamente un objetivo ni una estrategia relativa a la recogida de información (es el caso del “vamos a hacer esto ... y a ver lo que sale”). También se trata de pseudoevaluación cuando se realiza el proceso inverso, decidiendo de antemano “qué es lo que quiero que salga”, y dedicando después todos los esfuerzos a justificar con la evaluación lo que va a coincidir con lo que ya había prefijado (cuando al final de un período laboral se prepara una memoria de lo efectuado, y se presenta como evaluación) (Stufflebeam y Webster, 1980).
- b) *Identificación de la medida con la evaluación*. La medida lo es de los datos obtenidos, y se incurre en este error cuando se reduce la evaluación a una mera cuantificación de datos, incluso si no es reducible a cifras la realidad que pretendemos evaluar. No se pueden absolutizar los datos, porque, en sí mismos, son insuficientes sin una referencia al contexto que les da sentido y relieve. Así, el número de jóvenes que ha participado en una determinada actividad de un barrio no es evaluable sin una alusión a la población total de jóvenes existentes en dicho barrio, o sin saber a qué hora y en qué fecha se realizó, o si es la primera vez o se trata de una actividad habitual. Se requiere completar los datos cuantitativos con referencias cualitativas y contextuales para poder valorarlos adecuadamente.
- c) *Sectarismo o parcialidad*. Suele aparecer debido a los prejuicios, intereses o apasionamiento que frecuentemente acompañan a las evaluaciones. Puede haber tendenciosidad tanto en la selección y obtención de informaciones como en el momento de difundirla.
- d) *Subjetivismo*. Afecta a la evaluación cuando los datos no están convenientemente fundamentados o contextualizados.
- e) *Confusión de niveles*. Es necesario iniciar la evaluación sabiendo *qué, para qué y a quién* le interesa esa evaluación. En función de la respuesta que se dé a cada una de estas preguntas se establecen distintos niveles e intereses, los cuales no pueden mezclarse. Así, a la institución que patrocina una determinada actividad comunitaria le puede interesar la relación coste/beneficio por participante, mientras que al usuario le preocupa el grado de cumplimiento de sus expectativas, y el profesional puede estar interesado en la calidad del programa. En cada evaluación existen distintos y hasta contradictorios intereses, y el evaluador tiene que dar a cada uno lo suyo, sin confundir la perspectiva institucional con la comunitaria o la profesional.
- f) *Burocratización*. Es muy frecuente cuando la evaluación se realiza en instituciones muy normadas, en las que la rutina formal y los trámites administrativos llegan a actuar a modo de barreras que retrasan y obstaculizan el proceso de evaluación.

C. Incidencia de la política en la evaluación.

Todas estas posiciones tienen en cuenta, en mayor o menor grado, el papel de la política en evaluación de programas.

Es indudable que hay programas políticamente legitimados (Cook, 1988) y otros no, independientemente del país que se considere, de la misma forma que “la evaluación de programas emerge como un ámbito de investigación y teorización dentro de las ciencias sociales *inseparablemente* ligado a la actuación política por lo que no tiene nada de particular que una de las principales polémicas se haya centrado en si se trata, esencialmente, de un ámbito científico o de una acción política” (Fernández-Ballesteros, 1995, p. 40).

Según Nicolau (1996) el responsable político y el agente técnico han de dar respuesta a un conjunto de problemas en los cuales la frontera entre técnico y político no es clara, dados los riesgos de vacío (porque hay papeles que nadie asume), confusión (porque hay papeles que son asumidos de forma concurrente por un político y un técnico), y de caos (porque hay papeles que son asumidos por más de un político o por más de un técnico). Por este motivo propone partir de un marco de trabajo acordado entre las partes (política y técnica) que sirva de marco de referencia.

Gran parte de los programas o intervenciones -en temas tales como, por ejemplo, la salud mental, la asistencia social, la educación, o planes de desarrollo- son diseñados partiendo de unos objetivos planteados en los programas políticos y referidos a ideologías concretas (Fernández-Ballesteros, 1983), o al menos el político habrá intervenido en su priorización (Nicolau, 1996) en la asignación de recursos, aunque aquí no nos interese desde el abordaje metodológico al que nos dirigimos, ya que el proceso científico que implica la evaluación corresponde al evaluador profesional, y no, evidentemente, al político. Es decir, una vez establecidas las prioridades, usualmente de orden político, tendremos que preocuparnos por implementar de la forma más rigurosa posible el proceso evaluativo.

D. Funciones sociales de la evaluación.

Finalmente, tenemos que contemplar la función social de la evaluación (Casas, 1996; Fernández del Valle, 1996b), en tanto en cuanto los usuarios de un programa se espera que se beneficien con el logro de los objetivos fijados. Otra cuestión es si los objetivos son los que más les convienen, o los que más desean los usuarios o sus allegados, o si se corresponden con sus necesidades percibidas, o si se cumplen las expectativas de los potenciales usuarios. Son cuestiones que todavía no están perfectamente delimitadas en la mayoría de evaluaciones realizadas.

1.6. Referencias

- Achúcarro, C. y San Juan, C. (1996). Aspectos históricos, teóricos y metodológicos de la evaluación de programas. En C. San Juan (Ed.) *Intervención psicosocial. Elementos de programación y evaluación socialmente eficaces* (pp. 56-89). Barcelona: Anthropos.
- Aguilar, M.J. y Ander-Egg, E. (1992). *Evaluación de servicios y programas sociales*. Madrid: Siglo XXI.
- Alkin, M.C., Patton, M.Q., Weiss, C.H., Conner, R., House, E., Kean, M.H., King, J., Klein, S., Law, A. y McLaughlin, M.W. (1990). *Debates on evaluation*. Newbury Park: Sage.
- Alvira, F. (1985). La investigación evaluativa: Una perspectiva experimentalista. *Revista Española de Investigaciones Sociológicas*, 29.
- Alvira, F. (1991). *Metodología de la evaluación de programas*. Madrid: Centro de Investigaciones Sociológicas.
- Anguera, M.T. (1989). Innovaciones en la metodología de la evaluación de programas. *Anales de Psicología*, 5 (1-2), 13-42.
- Anguera, M.T. (1990). Programas de intervención. ¿Hasta qué punto es factible su evaluación? *Revista de Investigación Educativa*, 8 (16), 77-93.
- Anguera, M.T. y Blanco, A. (1988-91). *Evaluación de programas en Servicios Sociales: Alternativas metodológicas*. Proyecto de investigación subvencionado por la Comisión Interministerial de Ciencia y Tecnología (CICYT). Barcelona: Universidad de Barcelona.
- Anguera, M.T., Blanco, A., Gil, A., Ibern, I., Losada, J.L., Quera, V., Redondo, S. y Roca, M. (1989-90). *Instrumentos y técnicas de evaluación de programas de rehabilitación en el ámbito penitenciario*. Barcelona: Consejería de Justicia de la Generalidad de Cataluña, Dirección General de Servicios Penitenciarios y de Rehabilitación.
- Anguera, M.T., Blanco, A., Losada, J.L., Sánchez Algarra, P., Hernández Mendo, A., Chacón, S., Ardá, A., Camerino, O. y Castellano, J. (2001-04). *Innovaciones en la evaluación de contextos naturales: Aplicaciones al ámbito del deporte*. Proyecto de investigación [BSO2001-3368] subvencionado por el Ministerio de Ciencia y Tecnología.
- Anguera, M.T. y Chacón, S. (1999). Dimensiones estructurales de diseño para la evaluación de programas. *Apuntes de Psicología*, 17 (3), 175-192.
- Anguera, M.T. y Redondo, S. (1991). La evaluación de la intervención penitenciaria. *Delincuencia / Delinquency*, 3 (3), 245-289
- Artells, J.J. (1989). *Aplicación del análisis coste-beneficio en la planificación de los servicios sanitarios. Eficiencia y equidad en la atención perinatal*. Barcelona: Masson.
- Ato, M., Quiñones, E., Romero, A. y Rabadán, R. (1989). Evaluación de programas. Aspectos básicos. *Anales de Psicología*, 5 (1-2), 1-12.

- Bartolomé, M. (1990). Evaluación y optimización de los diseños de intervención. *Revista de Investigación Educativa*, 8 (16), 39-59.
- Bell, W. (1983). *Contemporary social welfare*. Nueva York: MacMillan.
- Berk, R. y Rossi, P.J. (1990). *Thinking about program evaluation*. Beverly Hills: Sage.
- Berk, R. & Schulman, D. (1995). Public perceptions of global warming. *Climatic change*, 29, 1-33.
- Bowman, M.L. (1989). Testing individual differences in ancient China. *American Psychologist*, 44, 576-578.
- Blanco Villaseñor, A. (1992). Aplicaciones de la teoría de la generalizabilidad en la selección de diseños evaluativos. *Bordón*, 43 (4), 431-459.
- Brown, J.L. & Kiernan, N.E. (2001). Assessing the subsequent effect of a formative evaluation on a program. *Evaluation and Program Planning*, 24 (2), 129-143.
- Cabrera, F. (1987). La investigación evaluativa en educación. En *Técnicas de evaluación y seguimiento de programas de formación profesional*. Barcelona: Fundación Largo Caballero.
- Campbell, D.T. (1969). Reforms as experiments. *American Psychologist*, 24, 409-429.
- Campbell, D.T. (1970). Considering the case against the experimental evaluations of social innovations. *Administrative Science Quarterly*, 15, 110-113.
- Campbell, D.T. (1963). From description to experimentation: Interpreting trends as quasiexperiments. En C.W. Harris (Ed.) *Problems of measuring change* (pp. 212-242). Madison, Wisconsin: University of Wisconsin Press.
- Campbell, D. y Stanley, J. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally Company.
- Cabrera, F. (1987). La investigación evaluativa en educación. En *Técnicas de evaluación y seguimiento de programas de formación profesional*. Barcelona: Fundación Largo Caballero.
- Carballo, R. (1992). El impacto de los estudios evaluativos. *Bordón*, 43 (4), 509-515.
- Caride, J.A. (1989). *Diseño y evaluación de programas sociales*. Comunicación presentada al I Congreso de Animación Sociocomunitaria. Salamanca.
- Caride, J.A. (s/f). *De la evaluación de necesidades a la evaluación de programas sociales en desarrollo comunitario*. Documento interno. Santiago de Compostela: Universidad de Santiago de Compostela.
- Casas, F. (1996). *Bienestar social. Una aproximación psicosociológica*. Barcelona: P.P.U.
- Chacón, S., Anguera, M.T., Pérez-Gil, J.A. & Holgado, F.P. (2002). A Mutual Catalytic Model of Formative Evaluation: The Interdependent Roles of Evaluators and Local Programme Practitioners. *Evaluation. The International Journal of Theory, Research and Practice*, 8 (4), 413-432.
- Chacón, S., Carmona, J., Pérez-Gil, J.A., Holgado, F.P. y Losada, J.L. (2000-03). *Optimización del diseño en la evaluación de la transferencia al puesto de trabajo en programas de formación continua*. Proyecto de investigación [BSO2000-1462] subvencionado por el Ministerio de Ciencia y Tecnología.
- Chacón, S., Cook, T., Pérez-Gil, J.A., Anguera, M.T. & Shadish, W. (2001, September). Convergence between quasi-experimental and observational designs to enhance evaluative results generalization. *6th European Conference on Psychological Assessment*. Aachen (Alemania).

- Chacón, S., y J. López Ruiz (1993). La metodología de la evaluación de programas de intervención: Una aplicación en centros asistenciales infantiles. *Apuntes de Psicología*, 37, 41-60.
- Chacón, S., y J. López Ruiz (1995). Metodología de la evaluación de programas de intervención: Una aplicación en centros asistenciales infantiles II. *Apuntes de Psicología*, 45, 69-89.
- Chacón, S. y López Ruiz, J. (1998). Validez de la evaluación en intervención psicosocial: Un análisis metodológico. *Intervención Psicosocial*, 7 (3), 407-425.
- Chacón, S., Pérez-Gil, J.A. & Holgado, F.P. (2001). Evaluation of intervention programs for the elderly: Enhancing validity. In J. León-Carrión & M. Giannini (Eds.), *Behavioral neurology in the elderly* (pp. 214-253). Florida: CRC Press.
- Chacón, S., Pérez-Gil, J.A., Holgado, F.P., Lara, A., López, J.M. y Sanduvete, S. (1998-03). Sistemas de evaluación de los programas de formación de la Universidad de Sevilla. Proyecto de investigación subvencionado por el Secretariado de Formación Permanente de la Universidad de Sevilla. Regulado por los Contratos 11.45 de la L.R.U. y los Arts. 68 y 34 de la L.O.U., autorizado por el Vicerrectorado de Investigación de la Universidad de Sevilla.
- Chacón, S., Pérez-Gil, J.A., Holgado, F.P. y Lara, A. (2001). Evaluación de la calidad universitaria: Validez de contenido. *Psicothema*, 13, 294-301.
- Chacón, S. & Shadish, W. (2001). Observational studies and quasiexperimental designs: Similarities, differences and generalization. *Metodología de las Ciencias del Comportamiento*, 3 (2), 283-290.
- Chambers, F. (1994). Removing confusion about formative and summative evaluation: Purpose versus time. *Evaluation and Program Planning*, 17 (1), 9-12.
- Chen, H. y Rossi, P.H. (1983). Evaluating with sense: The Theory-driven approach. *Evaluation Review*, 7, 283-302.
- Cohen, E. y Franco, R. (1992). *Evaluación de proyectos sociales*. Buenos Aires: Grupo Editor Latinoamericano.
- Cook, T.D. (1988). Theories of program evaluation: A short history. *Evaluación Psicológica / Psychological Assessment*, 4 (1), 3-29.
- Conner, R.F. (1984). The art of implementation. En L. Rutman (Ed.), *Evaluation research methods: a basic guide* (pp.191-218). Beverly Hills, California: Sage.
- Conner, R.F., Altman, D.G, y Jackson, C. (1984). *Evaluation Studies. Review Annual*, 9, 26-48.
- Cochrane, A.L. (1985). *Eficacia y eficiencia. Reflexiones al azar sobre los servicios sanitarios*. Barcelona: Salvat.
- Crane, J.A. (1988). Evaluation as scientific research. *Evaluation Review*, 12(5), 467-482.
- Cronbach, L.J. (1963). Course improvement through evaluation. *Teachers College Record*, 64, 672-683.
- Cronbach, L.J. (1980). Validity on parole: How can we go straight? En Schrader, W.B. (Ed.) *New directions for testing and measurement, planning achievement: Progress over a decade*. San Francisco: Jossey-Bass.
- Cronbach, L.J. (1981). *Toward a reform of program evaluation*. San Francisco: Jossey-Bass.
- Cronbach, L.J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.

- Dehar, M., Casswell, S. & Duignam, P. (1993). Formative and process evaluation of health promotion and disease prevention programs. *Evaluation Review*, 12 (4), 449-461.
- De la Orden, A. (1990). Evaluación de los efectos de los programas de intervención. *Revista de Investigación Educativa*, 8 (16), 61-76.
- Donabedian, A. (1980). *The definition of quality and approaches to its assessment*, 2 vols. Ann Arbor, Michigan: Health Administration Press.
- Echebarría, A. y San Juan, C. (1996). Intervención psicosocial en el entorno hospitalario. Intervención psicológica en situaciones quirúrgicas. En C. San Juan (Coord.) *Intervención psicosocial. Elementos de programación y evaluación socialmente eficaces* (pp. 111-124). Barcelona: Anthropos.
- Espinoza, M. (1986). *Evaluación de proyectos sociales*. Buenos Aires: Humanitas.
- Esteve, R., Godoy, A., Rodríguez-Naranjo, C. & Fierro, A. (1993). Program evaluation use: An empirical study. *European Journal of Psychological Assessment*, 9 (2), 77-93.
- Fernández-Ballesteros, R. (1983, 30 Marzo). *La evaluación de programas sociales*. El País.
- Fernández-Ballesteros, R. (1984). Evaluación de programas para el desarrollo de la inteligencia. En UNESCO, *Evaluación de programas para el desarrollo de la inteligencia implantado en Venezuela por el MEDI*. Paris: UNESCO.
- Fernández-Ballesteros, R. (1985). Evaluación psicológica y evaluación valorativa. *Evaluación Psicológica / Psychological Assessment*, 1-2, 5-24.
- Fernández-Ballesteros, R. (1986). *Evaluación de programas del INSERSO*. Memoria de investigación (Documento no publicado). Madrid: Universidad Autónoma de Madrid.
- Fernández-Ballesteros, R. (1987, Septiembre). Ciencia, ideología y política en evaluación de programas. *Revista de Psicología Social*, 2, 159-183.
- Fernández-Ballesteros, R. (1988). *Listado de cuestiones relevantes en evaluación de programas (LCREP)*. Manuscrito no publicado. Madrid: Universidad Autónoma, Facultad de Psicología.
- Fernández-Ballesteros, R. (1992a). Valoración de programas. En R. Fernández-Ballesteros (Ed.) *Introducción a la Evaluación Psicológica, Vol. II* (pp. 473-503). Madrid: Pirámide.
- Fernández-Ballesteros, R. (1992b). A model for planning evaluation research. In J. Mayne, M.L. Bemelmans-Videc, J. Hudson & R. Conner (Ed.) *Advancing Public Policy Evaluation: Learning from International Experiences* (pp. 205-213). Amsterdam: Elsevier.
- Fernández-Ballesteros, R. (Ed.) (1995). *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.
- Fernández-Ballesteros, R. y Hernández, J.M. (1995). Listado de cuestiones relevantes en evaluación de programas (L.C.R.E.P.). En R. Fernández-Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 339-352). Madrid: Síntesis.
- Fernández Ballesteros, R., Díaz Veiga, P., Izal, M. y González, J.L. (1987). Evaluación de una residencia de ancianos y valoración de intervenciones ambientales. En R.

- Fernández Ballesteros (Coord.) *El ambiente. Análisis psicológico* (pp. 227-249). Madrid: Pirámide.
- Fernández-Ballesteros, R., Izal, M., Hernández, J.M., Montorio, I. y Llorente, G. (1989). *Evaluación de programas del INSERSO*. Informe de investigación UAM-INSERSO. Madrid: Universidad Autónoma de Madrid.
- Fernández del Valle, J. (1996a). Introducción: Evaluación de programas. *Intervención Psicosocial*, V (14), 7-8.
- Fernández del Valle, J. (1996b). Roles y estrategias en evaluación de programas. *Intervención Psicosocial*, V (14), 9-21.
- Fitzpatrick, J. (1988). Roles of the evaluator in innovative programs. *Evaluation Review*, 12 (4), 449-461.
- Flay, B.R. (1986). Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Preventive Medicine*, 15, 451-474.
- Folkman, D.V. & Rai, K. (1997). Reflections on facilitating a participatory community self-evaluation. *Evaluation and Program Planning*, 20 (4), 455-465.
- Freeman, H.E. y Solomon, M. (1981). Introduction and the uncertain '80's. En H. Freeman y M. Solomon (Eds.) *Evaluation Studies. Review Annual*. Beverly Hills: Sage.
- García Ramos, J.M. (1992). Recursos metodológicos en la evaluación de programas. *Bordón*, 43 (4), 461-476.
- Garner, B.R., Godley, S.H. & Funk, R.R. (2002). Evaluating admission alternatives in an outpatient substance abuse treatment program for adolescents. *Evaluation and Program Planning*, 25 (3), 287-294.
- Gibbs, D., Napp, D., Jolly, D. Westover, B. & Uhl, G. (2002). Increasing evaluation capacity within community-based HIV prevention programs. *Evaluation and Program Planning*, 25 (3), 261-269.
- Greene, J.C. (1990). Technical quality versus user responsiveness in evaluation practice. *Evaluation and Program Planning*, 13, 267-274.
- Gregory, A. (2000). Problematizing participation. A critical review of approaches to participation in evaluation theory. *Evaluation. The International Journal of Theory, Research and Practice*, 6 (2), 179-199.
- Guba, E.G. (1987). Naturalistic evaluation. En D.S. Cordray, H.S. Bloom y R.J. Light. (Eds.) *Evaluation practice in review* (pp.23-44). San Francisco: Jossey-Bass.
- Herman, J.L; Morris, L.L. y Fitz-Gibbon, C.T. (Eds.). (1987). *Program evaluation kit* (2ª Edic.). Beverly Hills: Sage Publications. (9 Vols.).
- Hernández, J.M. y Rubio, V.J. (1992). El análisis de la evaluabilidad: Paso previo en la evaluación de programas. *Bordón*, 43 (4), 397-405.
- Hofstee, W.K.B. (1992). From the citizen's point of view: The positioning of program evaluation. En J. Mayne, M.L. Bemelmans-Videc, J. Hudson y R. Conner (Ed.) *Advancing public policy evaluation: Learning from international experiences* (pp. 277-283). Amsterdam North-Holland: Elsevier Science Publishers B.V.
- Holland, P.W. & Rubin, D.B. (1998). Causal inference in retrospective studies. *Evaluation Review*, 12 (3), 203-231.
- House, E.R. (1984). How we think about evaluation. En R.F. Conner, D.G. Altman y C. Jackson (Eds.) *Evaluation Studies Review Annual*, vol. 9 (pp. 81-101). Beverly Hills: Sage.

- Ibern, I. y Anguera, M.T. (1990) Evaluación de la viabilidad de los programas de intervención educativa. *Revista de Investigación Educativa*, 8 (16), 461-465.
- Johnson, R.B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21 (1), 93-110.
- Joint Committee on Standards for Educational Evaluation (1981, 1994). *Standards for evaluations of educational programs, projects, and materials*. Nueva York: McGraw-Hill.
- Keinonen, T. (1998). *Dimensions of usability. One-dimensional usability-influence of usability on consumers' product*. Helsinki: UIAH.
- Krumholz, N. (1982). A retrospective view of equity planning. *American Planning Association Journal*, 48, 163-174.
- Levin, R. (1999). Participatory evaluation: Researchers and service providers as collaborators versus adversaries. *Violence Against Women*, 5 (10), 1213-1227.
- Leviton, L.C., Collins, C.B., Laird, B.L. & Krat, P.P. (1998). Teaching evaluation using evaluability assessment. *Evaluation. The International Journal of Theory, Research and Practice* 4 (4), 389-409.
- Leviton, L.C. & Valdiserri, R.O. (1990). Evaluating AIDS prevention: Outcome, implementation, and mediating variables. *Evaluation and Program Planning*, 13 (1), 55-66.
- Linder, S.H. & Peters, B.G. (1990). Policy formulation and the challenge of conscious design. *Evaluation and Program Planning*, 13 (3), 303-311.
- López Blasco, A. (1992). Proyectos de evaluación de la praxis en Servicios Sociales. En *Simposium Nacional de Evaluación en Servicios Sociales* (pp. 9-31). Jaén: Diputación Provincial de Jaén.
- Lucky, J., Broughton, A. & Sorensen, J. (1983). Archival data in program evaluation and policy analysis. *Evaluation and Program Planning*, 5, 319-326.
- Luque, O. y Casquete, A. (1989). La evaluación de programas de intervención social: Criterios y técnicas. *Revista de Treball Social*, 115.
- Lysaght, R.M. & Altschuld, J.W. (2000). Beyond initial certification: The assessment and maintenance of competency in professions. *Evaluation and Program Planning*, 23 (1), 95-104.
- Madaus, G.F., Stufflebeam, D.L. y Scriven, M.S. (1983). Program evaluation: A historical overview. En G.F Madaus, M.S. Scriven y D.L. Stufflebeam (Eds.) *Evaluation models. View points on education and human services* (pp.3-23). Boston: kluwer-Nijhoff Publishing.
- Mathie, A. & Greene, J.C. (1997). Stakeholder participation in evaluation : How important is diversity ? *Evaluation and Program Planning*, 20 (3), 279-285.
- Mateo, J. (1990). La toma de decisiones en el contexto de la evaluación educativa. *Revista de Investigación Educativa*, 8 (16), 95-112.
- Medina, M.E. (1996). *Gestión de servicios sociales*. Madrid: DM.
- Medina, A. y Villar, L.M. (Eds.) (1995). *Evaluación de programas educativos, centros y profesores*. Madrid: Universitas.
- Modarresi, S., Newman, D.L. & Abolafia, M.Y. (2001). Academic evaluators versus practitioners: Alternative experiences of professionalism. *Evaluation and Program Planning*, 24, 1-11.
- Mohr, L.B. (1992). *Impact analysis for program evaluation*. Newbury Park: Sage.

- Municio, P. (1992). La evaluación segmentada de los programas. *Bordón*, 43 (4), 375-395.
- Muñoz Tortosa, J. (1992). Metodología de intervención en el seguimiento y evaluación de los programas de ayuda social a la infancia. En *Simposium Nacional de Evaluación en Servicios Sociales* (pp. 101-125). Jaén: Diputación Provincial de Jaén.
- Nicolau, R. (1996). *El rol del político y del técnico en la gestión de servicios sociales*. Trabajo presentado a las IV Jornadas de Intervención Social del Colegio Oficial de Psicólogos de Madrid: Transformación social y compromiso de los profesionales, Vol. 2 (pp. 1219-1230). Madrid: Ministerio de Asuntos Sociales.
- OMS (Organización Mundial de la Salud) (1981). *Evaluación de los programas de salud. Normas fundamentales*. Ginebra: OMS, Serie "Salud para todos", nº 6.
- ONU (Organización de Naciones Unidas) (1984). *Pautas básicas para el seguimiento y evaluación de programas*. Nueva York: Naciones Unidas.
- Pandiani, J.A., Banks, S.M., Bramley, J., Pomeroy, S. & Simon, M. (2002). Measuring access to mental health care: A multi-indicator approach to program evaluation. *Evaluation and Program Planning*, 25 (3), 271-285.
- Patton, M.Q. (1978). *Utilization-focused evaluation*. Beverly Hills: Sage.
- Pérez Serrano, G. (1993). *Elaboración de Proyectos Sociales*. Madrid: Narcea.
- Pineault, R. y Daveluy, C. (1987). *La planificación sanitaria. Conceptos, métodos, estrategias*. Barcelona: Masson.
- Posavac, E.J. y Carey, R.G. (1985). *Program evaluation. Methods and case studies*. Englewood Cliffs, N.Y.: Prentice-Hall.
- Qvortrup, M. (1992). Models in evaluation: A method to increase utilization. En J. Mayne, M.L. Bemelmans-Videc, J. Hudson y R. Conner (Ed.) *Advancing Public Policy Evaluation: Learning from International Experiences* (pp. 269-273). Amsterdam: Elsevier.
- Ramírez, J. (1992). La retórica como lógica de la evaluación. *Bordón*, 43 (4), 407-420.
- Robinson, E.J. y Whitfield, M.J. (1987). Participation of patients during general practice consultations. *Psychology & Health. An International Journal*, 1 (2), 123-132.
- Reboloso, E. (1994). La evaluación del personal profesional: ¿Un modelo de evaluación nuevo o una síntesis de modelos previos? *Revista de Psicología Social*, 9 (2), 231-255.
- Reboloso, E. (1995). Theoretical and methodological considerations in the evaluation of patients leaving the emergency room without receiving medical attention. *European Journal of Psychological Assessment*, 11 (1), 29-40.
- Redondo, S. (1993). *Evaluar e intervenir en prisiones. Un enfoque conductual aplicado*. Barcelona: P.P.U.
- Rossi, P.H. y Freeman, H.E. (1989). *Evaluación. Un enfoque sistemático para programas sociales*. México: Trillas.
- Rutman, L. (1977). *Evaluation research method: A basic guide*. Beverly Hills: Sage.
- San Juan, C., García, L., Sota, E. y Guimón, J. (1996). Intervención psicosocial en el entorno sanitario. Evaluación e intervención del síndrome del quemado en el personal asistencial. En C. San Juan (Coord.) *Intervención psicosocial. Elementos de programación y evaluación socialmente eficaces* (pp. 125-133). Barcelona: Anthropos.
- Sánchez Vidal, A. (1991). *Psicología comunitaria. Bases conceptuales y operativas. Métodos de intervención*. Barcelona: P.P.U.

- Scriven, M. (1967). The methodology of evaluation. En R.W. Tyler, R.M. Gagne y M. Scriven (Eds.) *Perspectives of curriculum evaluation*. Chicago: Rand McNally.
- Scriven, M. (1972). The methodology of evaluation. En C. Weiss (Ed.) *Evaluating action programs: Readings in social action and education*. Boston: Allyn & Bacon.
- Scriven, M. (1991). *Evaluation Thesaurus*. Newbury Park: Sage.
- Shadish, W.R., Cook, T.D. y Leviton, L.G. (1991). *Foundations of program evaluation*. Newbury Park: Sage.
- Shirattuddin, N. & Landoni, M. (2002). Evaluation of content activities in children's educational software. *Evaluation and Program Planning*, 25 (2), 175-182.
- Smith, E.R. y Tyler, R.W. (1942). *Appraising and recording student progress*. Nueva York: Harper y Row.
- Smith, M.F. (1990). Evaluability assessment: Reflections on the process. *Evaluation and Program Planning*, 13 (4), 359-364.
- Stake, R. (1975). *Evaluating the arts in education: A responsive approach*. Ohio: Merrill.
- Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher*, 7, 5-8.
- Stufflebeam, DL. y Shinkfield, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica*. Madrid: Paidós-MEC.
- Stufflebeam, D. y Webster, W. (1980). An Analysis of alternative approaches to evaluation. *Educational Evaluation and Policy Analysis*, 3 (2), 5-19.
- Suchman, E. (1967). *Evaluative research: Principles and practice in public service and social actions programs*. Nueva York: Sage.
- Tejedor, F.J. (1990). Perspectiva metodológica del diagnóstico y evaluación de necesidades en el ámbito educativo. *Revista de Investigación Educativa*, 8 (16), 15-37.
- Tessmer, M. (1994). Formative evaluation alternatives. *Performance Improvement Quarterly*, 7 (1), 3-18.
- Thompson, S.C. (1981). Will it hurt if I control it? A complex answer to a simple question. *Psychological Bulletin*, 90, 89-101.
- Torroba, I. (1992). Evaluar a través de la observación. *Bordón*, 43 (4), 495-507.
- Toulemonde, J. (1995). Should evaluation be freed from its causal links? An answer illustrated by European Economic Development Policies. *Evaluation and Program Planning*, 18 (2), 179-190.
- Ubel, P.A., DeKay, M.L., Baron, J. & Asch, D.A. (1996). Cost-effectiveness analysis in a setting of budget constraints. Is it equitable? *New England Journal of Medicine*, 334, 1174-1177.
- Van Vlaenderen, H. (2001). Evaluating development programs: Building joint activity. *Evaluation and Program Planning*, 24 (4), 343-352.
- Vedung, E. (1993). *Modelos de evaluación*. Revista de Servicios Sociales y Política Social, 2, 39-68.
- Vedung, E. (1995). Utilización de la evaluación de programas. En R. Fernández Ballesteros (Ed.) *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp.321-338). Madrid: Síntesis.
- Vega, F.J. y Ventosa, V.J. (1993). *Programar, acompañar, evaluar*. Madrid: CCS.

- Veney, J.E. y Kaluzny, A.D. (1984). *Evaluation and decision making for health services program*. Englewood Cliffs, N.Y.: Prentice-Hall.
- Weiss, C.H. (1975a). *Investigación evaluativa. Métodos para determinar la eficiencia de los programas de acción*. México: Trillas.
- Weiss, C.H. (1975b). Interviewing in evaluation research. In E.L. Struening & M. Guttentag (Eds.), *Handbook of evaluation research* (pp. 355-395). Beverly Hills: Sage.
- Wertheimer, M., Barclay, A.G., Cook, S.W., Kiesler, Ch.A., Koch, S., Riegel, K.F., Rorer, L.G., Senders, V.L., Smith, M.B. y Sperling, S.E. (1978). Psychology and the future. *American Psychologist*, 33 (7), 631-647.
- Whitmore, E. (1991). Evaluation and empowerment: It's the process that counts. *Empowerment and Family Support*, 2 (2), 1-7.
- Wholey, J.S. (1977). Evaluability assessment. En L. Rutman (Ed.) *Evaluation research methods: A basic guide*. Beverly Hills: Sage.
- Wholey, J.S. (1979). *Evaluation: Promise and performance*. Washington: Urban Institute.
- Wholey, J.S. (1989). Evaluability assessment: Developing program theory. En L. Bickman (Ed.) *Using program theory in evaluation* (pp. 77-92). San Francisco: Jossey-Bass.