

anales de psicología

Número Monográfico, 1989, 5, 13-42

© Secretariado de Publicaciones e Intercambio Científico.
Universidad de Murcia.

Innovaciones en la Metodología de Evaluación de Programas

*M. Teresa Anguera Argilaga*¹²

Universidad de Barcelona

¹Parte de este trabajo corresponde a la investigación *Evaluación de Programas en Servicios Sociales: Alternativas Metodológicas* (en colaboración con A. Blanco), subvencionada por la CICYT (1988-91).

²*Dirección del autor*: Departamento de Metodología de las Ciencias del Comportamiento. Universidad de Barcelona. 08028 BARCELONA.

anales de psicología

Número Monográfico, 1989, 5, 13-42

Abstract

In spite of its growth, the program evaluation has important methodological problems to solve. To make that possible, a previous elaboration of the framework is required as a corpus of knowledge in which the different paths to be followed are handled in a way constituting a complete and totally documented outline for any of the established situations.

Resumen

La evaluación de programas, a pesar de su desarrollo, presenta importantes lagunas a nivel metodológico. Por dicho motivo es preciso previamente elaborar el entramado del proceso a modo de corpus de conocimiento en que se articulen las diferentes vías a seguir de forma que constituya una pauta completa y específicamente documentada para cualesquiera de las situaciones planteadas.

Keywords

Methodological advances, design, program evaluation, assessment.

Palabras clave

Avances metodológicos, diseño, evaluación (de programas), evaluación (proceso).

1 Incidencia de las innovaciones metodológicas

La evaluación de programas se halla en pleno proceso de expansión, y hacemos nuestra la afirmación de Conner, Altman & Jackson (1984) en el sentido de que, como disciplina formal, se halla aún en la época de la adolescencia desde la mitad de los años setenta, a pesar de que en la década de los sesenta y principio de la siguiente se puede hablar de una auténtica "edad de oro" de la evaluación (Rossi & Wright, 1987).

Lo demuestra por una parte el hecho de su veloz crecimiento e interés despertado, traducido a nivel de publicaciones -incluso a nivel de monografías periódicas, como *Evaluation Studies Review Annual*, *New Directions for Program Evaluation*, o *Contemporary Evaluation Research*, o de revistas, como *Evaluación Psicológica / Psychological Assessment*, *Evaluation Review*, *Evaluation & Program Planning*, *Evaluation Practice*, *Evaluation & the Health Professions*, *Evaluation News*, etc.-, de la organización de Congresos de ámbito nacional, continental e internacional, de la impartición de programas de Doctorado específicos¹ a nivel académico y otros cursos de nivel superior (Davis, 1986), de la constitución de sociedades (Sociedad Española de Evaluación Psicológica, Evaluation Research Society, Evaluation Network, American Evaluation Society, Canadian Evaluation Society, etc.) y de la puesta en marcha de proyectos de gran envergadura respaldados por instituciones públicas y privadas de los países más desarrollados, fomentando incluso una colaboración internacional en cuestiones de investigación². Pero, por otra parte, no se puede afirmar que este crecimiento se realice de forma armónica ni que esté coordinado, y ni siquiera que exista una mínima base de acuerdo en el propio estatus de la evaluación de programas, materializándose esta heterogeneidad de posiciones en las polémicas sobre su carácter de ciencia básica vs. aplicada -y, en consecuencia, su grado de cientificidad (Campbell, 1982)-, la dicotomía evaluación de programas vs. investigación evaluativa, una prioridad más decantada al ámbito ideológico-sustantivo o al tecnológico, etc., lo cual redundo, evidentemente, en los recursos metodológicos utilizados y en la forma de abordarlos.

En definitiva, el propio concepto de **evaluación de programas** no es monolítico, y las diversas posturas adoptadas oscilan entre la construcción de un corpus de conocimientos que pudiera garantizarle el carácter de disciplina básica (Anguera y Blanco, 1988-91), y la mera recogida de información que trata de dar cuenta de la ejecución de ciertas actividades realizadas previamente, como un *servicio dedicado a* ... (Cordray, Bloom & Light, 1987), pasando por el análisis de los efectos netos de los programas o por la primacía de razones políticas -y no técnicas- respecto a la oportunidad de algunos programas. Existen, por tanto, múltiples puntos de vista, si no *paradigmas* (Cordray & Lipsey, 1987), acerca de cómo debe conducirse la evaluación,

¹ *Evaluación en las Ciencias del Comportamiento: Metodología y Aplicaciones* (Programa de Doctorado que se imparte en la Universidad de Barcelona), bienios 1987-89, 1988-90 y 1989-91.

² Próximamente (8-11 mayo 1990) tendrá lugar el II Ciccio (Italia) una reunión científica de representantes de los países encuadrados en la OTAN sobre *International Developments in Program Evaluation Research*.

con indudables implicaciones metodológicas, destacando, por ejemplo, la reforma de Cronbach, Ambron, Dornbusch, Hess, Hornik, Phillips, Walker & Weiner (1980), la evaluación consensuada (Gold, 1981), la evaluación naturalista (Guba & Lincoln, 1981), la evaluación focalizada (Patton, 1978), la evaluación de meta libre (Scriven, 1973), la evaluación responsiva (Stake, 1975), el modelo CIPP (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman & Provus, 1971), y la evaluación en base al adversario (Wolf, 1979), por citar sólo las perspectivas más conocidas.

Además, juega también un importante papel el carácter interdisciplinar de la evaluación de programas, en el que confluyen psicólogos, sociólogos, economistas, etc., lo que supone la adopción de posiciones muchas veces distintas que no siempre son fáciles de conjuntar en la elaboración, implementación, monitorización y análisis de los efectos de un programa.

De aquí que exista una indudable necesidad de estructuración en este ámbito, en el que deben fijarse las coordenadas básicas, para después incidir en aquellas cuestiones de carácter metodológico que sean susceptibles de constituir una innovación en evaluación de programas.

Partimos para ello de una conceptualización según la cual un programa se halla formado por un conjunto heterogéneo de acciones, a veces con una estructura difusa, cuyo objetivo es el de resolver algún problema que incumbe a los sujetos (usuarios o beneficiarios), y una vez se haya previamente fijado mediante un análisis de necesidades; en este mismo sentido, para Fernández Ballesteros (1987, p. 2) la evaluación de programas consiste en una *aplicación de procedimientos de investigación social para el enjuiciamiento o análisis sobre el mérito, valor o eficacia de acciones humanas concretas implantadas en la realidad social*. La evaluación de programas implica una cadena de tomas de decisión que se inicia en el interés de los objetivos en base a las necesidades³ detectadas (McKillip, 1987) y evaluadas (Neuber, 1980; Martí-Costa & Serrano-García, 1983), de acuerdo con alguno de los modelos existentes al efecto, y se concreta en una supuesta relevancia del impacto esperado de las actividades programadas, de forma que se produzca un cambio efectivo en la dirección deseada con el mínimo costo posible (Grabe, 1983).

Se trata, en definitiva, de un proceso que intenta determinar, tan estructurada y objetivamente como sea posible, la relevancia, efectividad e impacto de las acciones a la luz de estos objetivos (Sohm, 1978), no sin antes implicar sucesivas tomas de posición ante diversos tipos de dilemas (Schalock & Thornton, 1988). Precisamente la forma cómo se han presentado concatenadas las fases de que consta nos muestra la fineza de su desarrollo, cada vez mayor en la actualidad, en que se da rango de fase a cuestiones que antes se hallaban anidadas en otras, como la definición del problema (York, 1982; Mayer, 1985) y la implementación (Williams, 1975; Pancer & Westhues, 1989).

³A pesar de que el término "necesidad" se confunde en ocasiones con el de "demanda", sobre todo por parte de no especialistas, no especificamos aquí sus diferencias por entender que se trata de conceptos determinados socialmente y poco precisos, que desde un punto de vista analítico no tienen gran valor.

Con el fin de facilitar una sistematización relativa a innovaciones metodológicas (que toman como referencia -y por tanto no innovación o no uso de la misma- los más recientes trabajos publicados sobre **evaluación de programas**), se considerarán dos grandes apartados relativos al *proceso* y al *producto*, y, en consecuencia, las fases de elaboración e implementación del programa, por una parte, y la valoración del mismo, por otra.

2 Innovaciones en el proceso

2.1 Operativización de indicadores

Como consecuencia del análisis de necesidades y de la adecuación del modelo, una de las primeras decisiones de gran trascendencia será la elección de indicadores, y en este punto es cuando precisamente conviene distinguir entre los "teóricos" de la evaluación de programas y sus "gestionadores" (Neigher & Metlay, 1983; Clifford & Sherman, 1983). La labor de éstos no implica asumir funciones de carácter técnico-metodológico, y una inadecuada división de tales funciones halla en el establecimiento de indicadores uno de sus puntos más vulnerables.

La afirmación de Bunge (1975, p. 77) de que "en cualquier circunstancia, cualquier evaluación adecuada, objetivamente verdadera, de la calidad de vida, ha de incluir indicadores de variables psicológicas" plantea importantes problemas, dada la gran polémica suscitada en torno al propio concepto de indicador (Carmona, 1977; Casas, 1988). Desde que Bauer (1966) acuñara el concepto, se han elaborado varias decenas de definiciones, tendentes unas a enfatizar su carácter de medida estadística de un concepto o dimensión de éste, otros a considerarlo como componente de un modelo, y un último grupo a poner de manifiesto su carácter de definición operativa de un concepto. Y algo parecido ha ocurrido con su taxonomización.

Su papel en evaluación de programas es absolutamente fundamental, y es creciente su ritmo de utilización (Olsen & Mervin, 1977), salvo muy escasas excepciones (Sehldon & Freeman, 1970) en que se valoran negativamente, o en que se les acusa de "cuantificacionismo" (Carley, 1983). Se trata de instrumentos base que pueden detectar directa o indirectamente las modificaciones producidas, facilitando la determinación de aquellos aspectos en los que no es posible acceder a un suministro de datos de forma habitual.

Los indicadores se sitúan en la confluencia entre lo teórico y lo empírico en evaluación de programas, y mientras que la definición conceptual establece el significado en términos abstractos, los indicadores se refieren a sus características observables, medibles y empíricamente detectables (Sullivan & Feldman, 1979). El problema radica en aquellos casos en que determinados conceptos presentan un grado tal de abstracción que no son capaces de generar características observables, siendo necesaria una teoría

auxiliar para que puedan especificarse las relaciones entre indicadores y conceptos mediante un modelo o diagrama causal.

Una de las innovaciones en evaluación de programas que de aquí se derivan es la realización de un **análisis de variables latentes** (McCutcheon, 1987), sobre la premisa básica de que la covariación actualmente detectada entre las variables observadas se debe a las relaciones entre la variable manifiesta con la variable latente, siendo ésta la que explica las relaciones entre las observadas.

Además, en la elección de indicadores habrá que tener muy en cuenta la **validez**, o medida que revela efectivamente los cambios que se pretenden indicar; la **objetividad**, si produce los mismos resultados cuando se utiliza por diferentes sujetos; la **sensibilidad**, si es capaz de registrar los cambios de la situación o el fenómeno de que se trate; y la **especificidad**, si refleja los cambios exclusivamente en la situación estudiada. Así, la tasa de mortalidad en las instituciones penitenciarias (Anguera y Redondo, en prensa) es un indicador sensible, pero muy rudimentario, del estado de salud de la población reclusa; esta tasa es una medida directa de la mortalidad y sólo una medida indirecta del estado de salud, pero, sin embargo, es un indicador válido y objetivo, aunque no es específico respecto a ninguna medida sanitaria determinada debido a que la reducción de la tasa puede ser consecuencia de un elevado número de factores y difícilmente es atribuible a ninguna medida determinada.

2.2 Adecuación en el uso de recursos

En evaluación de programas se manejan recursos de diversa índole, que en parte se hallan en función de la necesidad que debe ser atendida, y en parte son independientes del objetivo específico del programa.

La elaboración de un programa debe ser fruto de técnicos, aunque la institución que lo respalda se halle inserta o mediatizada por intereses políticos (Palumbo, 1987). Precisamente la propia trayectoria seguida por los estudios sobre evaluación de programas nos muestra cómo las circunstancias políticas del lugar y momento influyeron y primaron determinados ámbitos, considerados más vulnerables, y casi siempre con el objetivo de lograr una mejora de la calidad de vida (Fernández Ballesteros, 1987; Rodríguez Castedo, 1987), sea de forma genérica, o bien atendiendo a colectivos que muestran unas mayores carencias.

Esto genera la necesidad de articular los recursos disponibles (Haggart, 1983) y organizarlos mediante una temporalización que permita obtener de ellos el máximo rendimiento (ver apartado 3.2.). En países como E.E.U.U. es frecuente hallar agencias dedicadas a elaborar inventarios de recursos (Petersen, Bosanac, Baranowski & Foren, 1981; Fitzgerald & Cutler, 1983; etc.), servicio que por otra parte es altamente apreciado por facilitar un intercambio o un uso compartido, o, al menos, por favorecer una búsqueda rápida de información; en España existen sólo algunos ejemplos (Saturno y Marset, 1980; Juan-Espinosa, Rubio y Márquez, 1987), siendo mucho más

atípico.

Un *primer criterio* de estructuración de los recursos en su ubicación debe ser por áreas geográficas u otras que faciliten una rápida comunicación interna (en España, sería por Comunidades Autónomas), y, a efectos organizativos, sería muy conveniente establecer subáreas (provincias, comarcas, ...). En cada una de estas zonas debe existir un responsable fácilmente localizable y disponerse de un mapa de programas clasificados sectorialmente (a partir de criterios homogéneos), y con indicación de

- Institución que lo respalda o avala (Administración Central, Comunidad Autónoma, Ayuntamiento, Organismo público o privado);
- necesidades concretas que trata de cubrir;
- colectividad a la que va dirigido;
- perfil profesional de las personas destinadas a la implementación;
- recursos mínimos necesarios ya poseídos y los adquiridos "ad hoc";
- posibilidad de reciclaje y nueva utilización de dichos recursos;
- fase en que se halla cada programa;
- problemas (si los hay) surgidos en su implementación,
- grado de sincronización/"decalage" con la temporalización llevada a cabo previamente y graficada mediante un Gantt, PERT (Program Evaluation and Review Technique) o CPM (Critical Path Method),
- progresos detectados desde que se inició la implementación del programa,
- análisis de costo/beneficio en la medida en que se disponga de los datos necesarios para ello. Asimismo, una vez que se consideren finalizados se dispondrá del correspondiente informe en el que se estudien y valoren sus efectos.

Este primer criterio es imprescindible para agilizar la evaluación de programas y facilitar una adecuación y complementariedad de los recursos. Sugerimos, además, la introducción de un *segundo criterio* relativo a la **capacidad** de los programas (número de ancianos que pueden disponer de cama en una residencia), **accesibilidad** (por ejemplo, posibilidad de que niños que viven en el extrarradio de grandes ciudades tengan cabida en un programa de escolarización) y **nivel de funcionamiento** (existen tablas de niveles que permiten objetivarlo, como los nueve fijados por el State of Illinois Department of Mental Health and Developmental Disabilities) en función de la densidad, intensidad y participación.

2.3 Diseños en evaluación

Durante el último cuarto de siglo la evaluación de programas se ha visto dominada por la metodología experimental, utilizada en el ámbito aplicado, especialmente en los campos social, de la salud, clínico, empresarial y educativo. Afirma el propio Campbell (1969) en un conocido artículo que para muchos científicos sociales de corte tecnocrático el experimento de campo aleatorizado prometía reemplazar el camino con traqueteos propio del ensayo y error por otro en que de forma consciente y racional la sociedad fuera sujeto y agente de experimentación.

Pero hace casi veinte años que vienen realizándose en evaluación de programas un número muy elevado de experimentos de campo y cuasi-experimentos, y ello acompañado, cada vez con mayor profusión, del uso de otras metodologías que no precisan del cumplimiento de requisitos de carácter restrictivo propios de la metodología experimental. Requisitos que, por otra parte, serían imposibles de mantener, dadas las características no pautadas de un tipo de estudio en que se incide sobre cuestiones y sujetos que, por formar parte de la vida misma, no pueden someterse a un control de todas las variables relevantes, y teniendo además en cuenta la habitual parte de varianza no explicada debido a variables extrañas que en muchas ocasiones sería prácticamente imposible dilucidar.

La cuestión es, sin embargo, más compleja. Se trata de un planteamiento "utilitarista" en tanto en cuanto interesa resolver unas carencias, en la mayor parte de los casos, que afectan a una comunidad de individuos (niños maltratados, ancianos que viven solos, adictos a drogas, inadecuada utilización del espacio urbano en un barrio, fracaso escolar, escasez de recursos humanos, inadecuadas redes de apoyo social, falta de adaptación social en presos que han cumplido la condena, etc.), y desde esta perspectiva es indudable el estudio de la *relación causa-efecto*, tanto por lo que se refiere a la solución del problema como también para una progresiva optimización del programa (sin que entremos ahora en detalles sobre modificaciones en su estructura o en su implementación). Pero aquí inciden fundamentalmente varios factores sumamente relevantes a nivel metodológico.

No abogamos por el criterio de manipulabilidad, pues aunque tradicionalmente ha servido de punto de referencia para la estructuración de las tres grandes metodologías (observacional, selectiva y experimental) (Arnau, Anguera y Gómez, en prensa), aquí no resulta ser el más adecuado por el simple motivo de que por tratarse de programas de intervención siempre existe en alguna medida manipulabilidad o control interno (Riecken & Boruch, 1974), y en consecuencia nunca se trata de acciones espontáneas, aunque, en contrapartida, debamos insertarnos en el estudio de situaciones complejas de la vida real. Por otra parte, y a pesar de que inicialmente, como se ha indicado, la evaluación de programas implicaba la elección de un diseño experimental o cuasi-experimental (Pancer & Westhues, 1989), con la restricción que éstos comportan para la búsqueda de relaciones de causalidad debido a los controles que deben establecerse (Lipsey, Crosse, Dunkle, Pollard & Stobart, 1985), no se considera ya admisible -al menos de forma absoluta- la aplicación de las pautas propias de la experimentación

(Anguera, Blanco, Gil, Ibern, Pérez, Quera y Redondo, 1989-90; Anguera y Redondo, en prensa) al ser conscientes del gran número de variables extrañas o perturbadoras (Alvira, 1989) que se hallan presentes (organísmicas, del sujeto, de funcionamiento institucional, de implementación de los programas, de funcionamiento comunitario, etc.) y las dificultades -muchas veces insalvables en su totalidad- que plantea su control mediante las técnicas habituales (mantenimiento de valores constantes, aleatorización, balanceo, contrabalanceo, etc.), como, por ejemplo, la asignación aleatoria de los tratamientos a los sujetos (Campbell & Boruch, 1975). De ahí que el análisis de la causalidad, si es factible, se lleve a cabo independientemente de la metodología base.

Además hay que tener muy presente, a nivel conceptual, la existencia de dos grandes modelos en evaluación (Veney & Kaluzny, 1984), el lineal y el no lineal, estrechamente relacionados con las denominadas evaluación sumativa y formativa, respectivamente, lo que trasciende totalmente sobre el diseño. El **modelo lineal** parte de la secuencialidad de las tres fases de planificación, implementación y evaluación de un programa, lo cual equivale a decir que la evaluación (medida de los efectos del programa) depende rígidamente de cómo se haya estructurado inicialmente un programa y de cómo se haya llevado a la práctica, no pudiendo ser moduladas estas dos grandes fases en función de circunstancias no previsibles que se derivan del carácter complejo que presenta todo estudio enraizado en la vida real. Por el contrario, el **modelo no lineal** se basa en la continua interacción y modulación recíproca entre planificación, implementación y evaluación, lo que equivale a la intersección continuada entre tres de los componentes del proceso de evaluación (relevancia, progreso y eficiencia, respectivamente), y esta flexibilidad nos lleva a optar claramente por dicho modelo (Anguera, en prensa a, en prensa b), a la vez que nos obliga a adaptar el diseño -ver nota 5 a pie de página- de forma que puedan introducirse posibles modificaciones sobre la marcha con el fin de ir optimizando los efectos.

A nivel práctico, se dificulta o incluso imposibilita el estudio de la simple relación causal, según los casos, y aunque este análisis siga efectuándose siempre que sea factible (Rubin, 1974; Curry, Yarnold, Bryant, Martin & Hughes, 1988; Chen, Quane, Garland & Marcin, 1988; etc.), pierde la primacía para dejar paso a otros -ver, por ejemplo, las 728 páginas del volumen 10 de *Evaluation Studies Review Annual* (1985) referido a la evaluación de programas sociales mediante estudios de campo-, de acuerdo con el objetivo y las distintas posibilidades de ordenación de los datos, y, por tanto, encaminado a una objetivación de las relaciones de concomitancia, covariación, secuenciación, etc., lo que constituye una base firme para la evaluación de las intervenciones programadas. No en vano la investigación evaluativa ha irrumpido con la fuerza que exige la demanda de su aplicación y se ha diversificado gracias a la multitud de enfoques existentes.

En este mismo sentido hay que tener muy en cuenta que nos hallamos insertos en un continuo proceso de toma de decisiones (Pitz & McKillip, 1984; Pollard, 1986) que constituye una parte irrenunciable de éste, ya que una vez se delimitan los programas, es decir, se establecen las metas y los medios (Artells, 1980), se toman las decisiones

encaminadas a la fijación de las acciones integradas que requieren, y se sigue su implementación -para lo que deberán seguir tomándose decisiones-, con una serie de servicios prestados y la subsiguiente recogida de información, decidiéndose de nuevo mediante los análisis adecuados la medida en que se alcanzan o no, y en qué grado, los efectos esperados por cada una de las acciones programadas y su conjunto.

En consecuencia, cabe plantear el diseño⁴ de evaluación en cada caso como la resultante de la confluencia de dos importantes criterios (Anguera y Redondo, en prensa), que son la *naturaleza de la respuesta elicitada* mediante las acciones del programa, y la *perspectiva temporal*, con la búsqueda de relaciones de causalidad, secuencialidad, concomitancia, ..., entre las acciones llevadas a cabo y los efectos que han generado.

2.3.1 Naturaleza de la respuesta elicitada

Se halla estrechamente vinculada a la recogida de datos, y a la cuestión ya planteada de los indicadores, y ha de adaptarse a la objetividad de una situación (Burstein & Freeman, 1985), que, por ser real, es indudablemente compleja y puede dar lugar a una distorsión de la información (Reiss, 1985), poniendo en juego para ello todas las estrategias que complementariamente se requieran, con o sin instrumento estandar para la obtención de la información. Este criterio, que ya venimos utilizando (Anguera y Redondo, en prensa), da lugar esquemáticamente a la Figura 1.

Cada una de las posibilidades expresadas da margen para múltiples variantes, especialmente cuando no existe pauta a seguir debido a la ausencia de instrumento estándar.

Cuando se dispone de *respuestas directamente perceptibles*, dado que únicamente interesa la obtención del registro de conducta, se aplicará la estrategia propia de la **observación directa como técnica**⁵, siendo objeto de estudio únicamente las conductas del ámbito al que se refiera el programa que comporten una percepción de la realidad. Por ejemplo, en programas motivacionales a presos, ¿cómo podría obtenerse información sobre conductas en el patio, el aula, la enfermería, la celda, el pasillo, etc., de la prisión?

Los principales problemas que se plantean se refieren a la fijación de las unidades de conducta, su sistematización y categorización, y la optimización del registro, así como la exploración de las diferentes posibilidades posteriores en el análisis de los datos,

⁴Utilizamos el término diseño desde su acepción no normativa, y por consiguiente no consideramos como la más adecuada la opción de mencionar los muchos diseños de carácter estandar que existen (véase Trochim, 1984; Veney & Kaluzny, 1984; Posavac & Carey, 1985, etc.), sino la resultante de plantear específicamente en cada caso cómo recoger, ordenar y analizar los datos.

⁵Téngase en cuenta que es distinto del uso de la metodología observacional, aunque a nivel mecánico coincidan las operaciones a llevar a cabo. Como *técnica* no significa más que una estrategia de recogida de un determinado tipo de datos subordinada a las directrices de una metodología selectiva o experimental (Arnau, Anguera y Gómez, en prensa). Aquí la observación tiene por misión únicamente el suministro de información de forma complementaria a las otras formas de recogida de datos y, por tanto, incluye el estudio de conductas generadas en situación artificial.

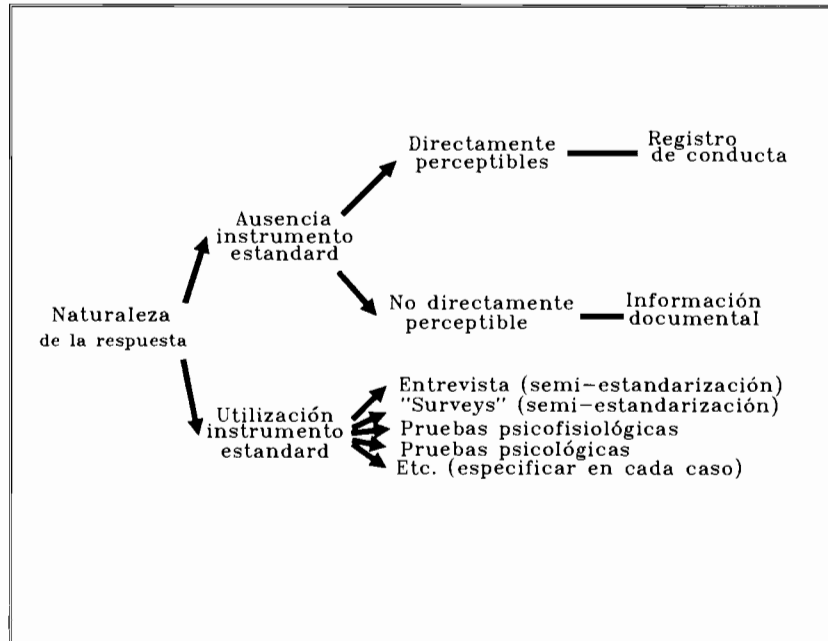


Figura 1:

pues aunque la información inicial es generalmente de naturaleza cualitativa, puede y debe ser tratada cuantitativamente una vez se ha llevado a cabo su codificación, lo cual comporta que la frecuencia, orden, duración, intensidad, etc. de las conductas registradas se contemplen en la métrica del registro.

Cuando obtenemos *respuestas no directamente perceptibles*, nos llevará a la forzosa aplicación de la estrategia propia de la observación indirecta como técnica. Esta implica la existencia de conductas encubiertas que requerirán una inferencia, y, por tanto, una carga interpretativa que puede redundar en detrimento y menoscabo de la objetividad requerida. Sin embargo, en los últimos años se ha avanzado enormemente en este sentido, y los materiales que forman parte de este grupo (indicadores externos de un estado emocional, textos documentales obtenidos por grabación de una entrevista abierta o por informaciones de terceros, o actas de reuniones, autoinformes, etc.) pueden ser sometidos a codificación con garantías de rigurosidad. En la evaluación de programas de Servicios Sociales son muchos los casos en que la información sólo puede obtenerse mediante entrevistas abiertas a los beneficiarios (por ejemplo, cuando se pregunta a los ancianos que viven solos en un barrio cómo se sienten respecto a compañía, ayuda, higiene, alimentación condimentada, etc. desde que acude periódicamente un asistente social a sus domicilios).

Si se cuenta con un *instrumento estandar de medida*, una vez solventado el problema de los indicadores, no existe ninguna otra dificultad a efectos de recogida de datos. Cabe destacar tan solo que en el caso de entrevistas pautadas y de "surveys" la

estandarización es parcial, y deberán resolverse cuestiones específicas de esta forma de recogida de datos -por ejemplo, el problema que se plantea ante la falta de respuesta (Rubin, 1987)-.

2.3.2 Perspectiva temporal

Las acciones que configuran un programa están entre sí lógicamente concatenadas, y no todas ellas tienen por qué surtir efecto simultáneamente, ni interesa por lo general saber si un programa resulta efectivo sólo en un determinado momento de tiempo.

La dimensión temporal en evaluación de programas tiene una importancia trascendental desde diversos ángulos:

- El efecto de un programa no tiene por qué ser puntual (salvo raras ocasiones, nunca se espera que lo sea), sino que debe extenderse en el tiempo, incorporando una **perspectiva diacrónica**, no siempre regular, que implicará un tratamiento adecuado de los datos en el diseño correspondiente, con lo que supone de seguimiento de una conducta, detección de los cambios producidos, aparición de tendencias, ciclicidad de determinados comportamientos, etc.

No nos vamos a referir a los clásicos diseños de estudio de un caso, de un solo grupo con pretest y postest y grupos de control no equivalentes que propusieron en su día Campbell y Stanley (1967), ni al diseño de discontinuidad en la regresión que Trochim (1984) nos presenta como propio de la investigación evaluativa por la estrategia seguida en la asignación de sujetos, ni tampoco a las series de tiempo - interrumpidas o no- de Cook & Campbell (1979), ni a otras alternativas que pudieran plantearse. Son tantas las posibilidades (Cook, 1981) en el planteamiento de los diseños de evaluación que no es posible fijar posiciones rígidas, sino dejar constancia de unas líneas básicas que sirvan de punto de referencia para el establecimiento de las pautas de actuación.

Queremos aquí únicamente llamar la atención desde la perspectiva diacrónica respecto a la *adecuación del diseño que se plantea en cada caso*, teniendo presente para ello el número de puntos de tiempo en que es factible la recogida de la información, la distancia entre ellos, y las condiciones que deben reunirse para que no se introduzcan sesgos, por otra parte frecuentísimos, al trabajar con sujetos que en muchos casos se hallan en colectividades a las que no siempre es fácil el acceso.

Un caso particular de diseño que está empezando a dar muy buenos resultados en evaluación de programas en el caso de contar con respuestas directamente perceptibles es el **diseño secuencial** resuelto mediante la técnica de retardos propuesta por Sackett (1978) y optimizada por él mismo (Sackett, 1980, 1987), que permite proceder a la obtención de patrones de conducta estandar, que después serán contrastados entre sí para analizar sus variaciones (Wampold & Holloway, 1983). Expresado con otros términos, equivale a obtener la configuración más o menos estable de conducta que subyace a un sujeto, haciéndola

"aflorar" a partir del registro, por lo que equivale a un extracto altamente condensado de la información obtenida, siendo muy útil para un conocimiento objetivo del comportamiento que interesa a efectos del programa aplicado, y pudiendo analizar su evolución, y, en consecuencia, el efecto del programa y la mayor o menor consolidación del cambio producido. Asimismo, es factible plantear inferencias causales a partir de los análisis de contingencia que pueden derivarse de la lista serial en el registro (Sackett, 1989).

Además, dado que en las conductas observadas en los sujetos previas y posteriores a la implementación del programa se pueden mantener entre sí distintos tipos de relación estadísticamente significativa (mútuamente excitatorias, mútuamente inhibitorias, y relación asimétrica entre una conducta -considerada focal- y cada una de las demás que componen el sistema de categorías en base al cual se efectuó el registro), es sumamente interesante analizar en sucesivas sesiones o grupos de ellas en períodos de tiempo consecutivos la medida en que se producen cambios en estas relaciones, dado el papel activador o aversivo de determinadas conductas y su posible aplicación a una remodelación del programa en el modelo no lineal. Este técnica, denominada "coordenadas polares" (Sackett, 1980), constituye otro ejemplo de gran adaptabilidad al referirnos al planteamiento de un diseño adecuado en cada caso y no normativo.

Del mismo modo, pero desde una **perspectiva sincrónica**, y especialmente en programas que constituyen sistemas complejos, interesa por lo general plantear un **diseño transversal** para conocer las relaciones de carácter simétrico y asimétrico (de contingencia y causalidad, respectivamente, y por consiguiente mediante modelos log-linear y logit) entre eventos registrados de forma cruzada, lo que equivale a decir entre sistemas alternativos de categorías (en base a criterios distintos) referidos a las conductas que interesan estudiar para comprobar los efectos del programa en un momento dado.

Por último, puede interesar combinar en una perspectiva mixta la diacrónica y la sincrónica si esta última se analiza no en un solo punto de tiempo, sino de forma continuada a lo largo del período establecido.

- Los diseños longitudinales que habitualmente nos vemos obligados a plantear en evaluación de programas son por lo general factibles de forma prospectiva (Boruch & Pearson, 1988), a pesar de que se presenten con gran frecuencia problemas que dificultan el seguimiento, como el de una mortalidad elevada de sujetos; pero la cuestión se complica cuando no queda otra alternativa que la de carácter retrospectivo (Holland & Rubin, 1988), en la que además se presenta agravado el problema de los "missing data" (Little & Rubin, 1987).

Los diseños longitudinales presentan indudables ventajas al ser accesibles las conductas bajo estudio -y por supuesto, incluyendo también las de carácter categórico (Landis & Koch, 1979; Plewis, 1985)- a lo largo de un período de tiempo prolongado, lo que permite analizar las variaciones producidas en su transcurso (Baltes & Nesselrode, 1979) y el estudio del cambio generado (Gold-

stein, 1979), aunque conviene dejar claro que no implican una metodología determinada, sino que todas ellas son factibles de uso en un estudio longitudinal. Puede llevarse a cabo un análisis de la causalidad en estudios longitudinales (Rogosa, 1979; Jöreskog, 1979; Plewis, 1985), así como hacer uso de las diversas variantes de éstos en el caso de que se adapten las exigencias y requisitos de cada uno a la naturaleza de nuestra información: técnicas multivariadas en general (Eber, 1983), análisis de panel (Kessler & Greenberg, 1981; Hsiao, 1986), de series de tiempo (Kratochwill, 1978; Frederiksen & Rotondo, 1979; Forehand, 1982), de series de tiempo interrumpidas (Gottman & Glass, 1978; McCain & McCleary, 1979), estudios de tendencias (Veney & Kaluzny, 1984), combinación de series de tiempo con diseños transversales (Mandell, 1989), etc., y teniendo en cuenta, además, que por lo general se dispone de escasos recursos para la puesta en marcha del programa, lo que revierte de nuevo sobre el diseño (Duncan, Juster & Morgan, 1984; Fienberg & Tanur, 1987).

3 Innovaciones en la valoración del producto

3.1 Previsión/estimación del impacto

La **previsión del impacto** está requiriéndose cada vez con mayor insistencia para un ajuste de plazos y el estudio de costos, y tiene lugar durante el proceso. Se trata de la actividad encaminada a identificar, predecir, interpretar, comunicar y prevenir el impacto de un programa, proyecto, plan o acción, y desemboca en la valoración de los efectos finales de una cadena de ellos que se inicia con una causa, que es precisamente la intervención o implementación del programa. Su origen se halla estrechamente vinculado al del *impacto ambiental*, aunque cada vez se halla más extendida la corriente que tiende a disociar los aspectos relativos a propiedades físicas y biológicas de los ambientes naturales de los aspectos pertenecientes a los ambientes humanos, encuadrando a éstos bajo la nomenclatura de *impacto social* (Finsterbusch, Llewellyn & Wolf, 1983; Rubio, 1987).

La previsión del impacto, en consecuencia, tiene un claro carácter prospectivo -lo que implica la posible discrepancia entre los valores predichos y los obtenidos-, con implicaciones conceptuales, metodológicas y de estrategia (Leistriz & Murdock, 1981), y se materializa a través de la secuencia propia de un proceso racional de toma de decisiones, que se inicia con la comprensión del problema (definir, enunciar, ..., diagnosticar la situación y explicitar los objetivos a alcanzar), se continúa con una fase de búsqueda de soluciones para encontrar cierto número de alternativas y termina en un juicio sobre el valor de cada una de ellas respecto a las metas propuestas y a la selección de la mejor, en lo que jugarán sin duda un papel relevante distintos criterios de economicidad (Domínguez, 1985).

El principal problema que conlleva es de carácter metodológico (Finsterbusch, 1985), e implica la búsqueda de relaciones funcionales y el estudio de la causalidad, que generalmente se asocia al efecto real del paso del tiempo (Anguera y Redondo, en prensa), y para lo que no falta un cierto componente de incertidumbre (Cordray & Sonnefeld, 1985), por otra parte lógica, a pesar de la existencia de los *Standards for Program Evaluation* (Rossi, 1982). Tanto la búsqueda de relaciones funcionales como el estudio de la causalidad deben plantearse con cautela, ya que no siempre son conocidos exhaustivamente ni los antecedentes ni los consecuentes, y en muchas ocasiones presentan importantes dificultades de operativización (Domínguez, 1985). Las técnicas más frecuentemente utilizadas son las listas de revisión (tipo Leopold, tipo Batelle, las propuestas por organismos nacionales o internacionales, etc.), cuestionarios para la identificación de aspectos conflictivos (emisión de juicios rápidos sobre proyectos), matrices de revisión causa-efecto para la búsqueda de relaciones funcionales (para cada una de las actividades seleccionadas en la intervención se revisan todos los factores ambientales/sociales/comportamentales que pueden verse afectados por su acción), matrices cruzadas (para el establecimiento de relaciones o efectos primarios, secundarios...), matrices escalonadas (representación matricial de la red), modelos de simulación, etc.

La **estimación del impacto** consiste en el análisis de los efectos del programa, y, por tanto, la medida en que la previsión del impacto se ajusta a la realidad. En un ejemplo hipotético en que se aplicara un diseño experimental de grupo control, la estimación del impacto sería la diferencia entre ambos grupos medida mediante el valor de la variable dependiente.

Obviamente el cálculo de dicha estimación vendrá determinado por el diseño planteado, y en el caso en que se lleve a cabo un seguimiento prolongado de los efectos del programa, la estimación podrá tener diferentes respuestas, tantas como momentos de tiempo se elijan para la medida de los efectos.

De acuerdo con los presupuestos del modelo no lineal de evaluación, los primeros efectos obtenidos podrán incidir en "feed-back" sobre la estructura del propio programa, o bien sobre la forma de llevar a cabo la implementación, tratando así de optimizar progresivamente el programa hasta que coincidan la previsión y la estimación del impacto.

Veney & Kaluzny (1984, p. 22) afirman textualmente que "la supervivencia de un programa debe ser considerada como una medida de eficiencia del programa", y ello nos debe llevar a plantear seriamente una nueva cuestión: ¿Cómo pueden estudiarse las diferencias entre programas a lo largo del tiempo? ¿Cómo repercuten de forma diferencial sobre los sujetos? ¿En qué medida la supervivencia de unos sujetos está en función de la "vida" de un programa? En el caso de programas sanitarios, por ejemplo contra el SIDA (Kaplan & Abramson, 1989) o las drogas (Fisher & Anglin, 1987; Anglin & Fisher, 1987), o de evitación de accidentes de tráfico (Maxim, 1989), o en general de aquellos que pretenden de forma directa preservar la vida de los ciudadanos, pueden darse unas circunstancias específicas, como el que los sujetos vayan entrando en el estudio a medida que estén disponibles, o que se consigan, o que

acepten voluntariamente en distintos momentos, y deban tratarse, en definitiva, muestras extendidas a lo largo del tiempo en forma irregular o no uniforme, no pudiéndose seguir los cánones habituales respecto a su obtención (Burnam & Koegel, 1988). En estos casos una interesante y sugerente posibilidad es el **análisis estadístico de supervivencia** (Cox & Oakes, 1984; Crowley & Breslow, 1984; Carrasco, 1986; Palmer, 1988) que permite, mucho mejor que las técnicas convencionales, hallar el patrón de los datos a lo largo del tiempo, y, por tanto, la probabilidad de "sobrevivir" (no reiniciendo) en cada instante temporal durante el período considerado, obteniéndose la correspondiente curva de supervivencia para cada programa, y mostrándose evidentemente sus diferencias. Finalmente, y muy relacionado con la previsión del impacto, por cuanto incide en la proyección de nuevos programas, están los estudios de **estimación de demandas futuras** (ver nota 3), en que se pretende analizar el comportamiento individual ante diferentes cuestiones (por ejemplo, consumo sanitario, lo que implicaría una demanda plasmada en número de visitas a médicos y estancia en hospitales y clínicas, y a su vez en función de variables de tipo demográfico, médico, socio-económico y de precios) para después poder predecir el comportamiento agregado (Galarraga, 1980). Generalmente se calcula la probabilidad esperada de que las personas pertenecientes a una determinada categoría previamente establecida tengan un comportamiento específico, y puede realizarse mediante un **modelo probit multivariante** (Aldrich & Nelson, 1984) en un caso como el mencionado, dada la naturaleza binaria del problema y el carácter aleatorio de la incidencia de la enfermedad, así como el hecho de acudir o no al médico. Pero, por supuesto, la técnica de análisis sería distinta ante otra expectativa, y así, por ejemplo, se trabajaría con un **modelo logit multivariante de respuestas multinómicas** (Aldrich & Nelson, 1984) si interesara conocer la probabilidad de que cada sujeto perteneciente a cada "categoría de planeamiento" (como renta, edad, sexo, ubicación, etc.) demandara un número determinado de visitas o actos médicos (0, 1, 2, 3, 4 ó más) en un período considerado, o se hospitalizara un número determinado de días o semanas.

3.2 Análisis de costo-beneficio

El análisis de costo-beneficio constituye en la actualidad una de las facetas que se halla en más rápido desarrollo en evaluación de programas. La cuestión fundamental estriba en la identificación de ambos montantes (costes y beneficios) y en cómo conseguir mayores y mejores recursos, tanto respecto al propio programa cuyo estudio interesa como a otros alternativos (Garfinkel & Corbett, 1985), incidiendo grandemente en la toma de decisiones respecto a primar la conveniencia de uno de ellos, y teniendo además en cuenta que la perspectiva es distinta a nivel individual, comunitario e institucional (Rossi & Freeman, 1985). Por supuesto, aunque en términos económicos tendría primacía aquel programa en que los beneficios superaran en mayor cuantía a los costos, no podemos en conciencia afirmar lo mismo, dado que existen muchos programas no lucrativos, considerados de interés social, en que se invierte

dicha relación.

Al margen de la diversificación del costo en las variantes de costo-eficiencia, costo-beneficio, costo-utilidad y costo-viabilidad, y que han supuesto avances metodológicos importantes en la última década, interesa destacar en especial las posibilidades del denominado **método de los ingredientes** (Lewin, 1983) para la estimación del costo.

En efecto, cualquier programa tiene un resultado o efectos y un costo. Así, en un programa de intervención penitenciaria (Anguera y Redondo, en prensa) el resultado es el efecto de la intervención, como, por ejemplo, mejor comportamiento de los internos en su relación con los funcionarios o con otros internos, disminución de autolesiones, aumento de asistencia a las clases de Graduado Escolar, menor consumo de drogas, más limpieza en la celda, etc. Pero cada intervención precisa de unos recursos, algunos de los cuales, como personal, locales, y otros, no son específicos para aquel programa de intervención, mientras que otros sí pueden serlo. Pues bien, el valor de los recursos que sí se emplean específicamente (se "gastan") representa el costo de una alternativa, lo cual, expresado en otros términos, equivale a lo que debemos perder por no utilizar de nuevo los recursos empleados. Técnicamente, entonces, el costo de una intervención específica se definiría como el valor de todos los recursos utilizados que hubieran sido asignados a su uso alternativo más valioso.

Precisamente el método de los ingredientes permite un acercamiento muy adecuado a esta cuestión, ya que, una vez fijados en función del correspondiente modelo teórico, o "teoría del programa" (Bickman, 1987; Conrad & Miller, 1987) -aunque hay que reconocer que es frecuente una gran ignorancia en esta cuestión por parte del equipo de evaluación (Cannon, McGuire & Dickey, 1985)-, se asigna un valor a cada uno de ellos, y con la adición de sus valores se establece el costo de la intervención, de forma que se pueda estudiar posteriormente cómo se parcelan o dividen dichos costos a efectos de quién lo subvenciona, lo que implica además la confección de un presupuesto, que en muchas ocasiones se revela como erróneo. Ahora bien, la asignación de valor a los ingredientes, en donde además deberá diferenciarse el momento temporal en que se crearon (Gramlich, 1981), ofrece serias dificultades cuando no existe precio de mercado, dando lugar a la *estimación de costos bajo incertidumbre* (Lewin, 1983; Mishan, 1988), tanto en el caso en que no existen estándares fiables que puedan tomarse como punto de referencia, como cuando se dispone de un rango o abanico de costos para un determinado ingrediente, y máxime si se trata de equipos o material cuya obtención ya no es viable. En este caso se impone un análisis de "sensitividad", propuesto por Stokey & Zeckhauser (1978), y que se apoya en una estimación de los costos bajo diferentes suposiciones para conocer la medida en que se altera el total y si se modifica el orden prioritario de las alternativas. Sugerimos en este punto, que es el menos desarrollado, la aplicación de la teoría de la decisión para resolver el problema.

Por lo que se refiere a los **beneficios**, en muchas ocasiones -las más- no son de carácter monetario (Nagel, 1983a, 1985), sino social, de bienestar, psicológico, político, etc., o de evitación de un gasto (Levey, Nyman & Haugaard, 1988), por lo que interesa enormemente la búsqueda de sistemas de "traducción" a términos económicos. De los

tres sistemas existentes (análisis incremental, método de porcentajes y comparaciones apareadas), el tercero, que por otra parte resulta ser el preferido, requiere aún de aportaciones metodológicas -teoría de la decisión, de nuevo-, ya que interesa que la *monetización* refleje los valores de las decisiones adoptadas incluyendo los de carácter subjetivo que caracterizan a los beneficios obtenidos.

Hay que tener en cuenta, sin embargo, que es más frecuente que lo deseado la existencia de factores muy difíciles o imposibles de detectar que juegan un papel importantísimo en la evaluación de programas por la medida en que contribuyen a su éxito o fracaso (Nagel, 1983b), y por tanto, inciden en costos o en beneficios. Por ejemplo, el cambio de director (que podrá ser más o menos simpático y atento que el anterior) en un centro de planificación familiar, a pesar de que se sigan implementando los mismos programas, va a ejercer indudablemente una influencia.

3.3 Meta-evaluación

Si nos hemos referido a innovaciones metodológicas en el proceso y en la valoración del producto, es lógico que incluyamos una referencia sobre la evaluación de la propia evaluación, además de constatar su necesidad para evitar vicios y sesgos que de lo contrario correrían el riesgo de consolidarse.

Son muy escasos los trabajos existentes sobre meta-evaluación en evaluación de programas (Stufflebeam, 1978; Cleave-Hogg & Byrne, 1988), y son precisamente los últimos avances en técnicas estadísticas de meta-análisis, éstas muy frecuentes en la literatura científica actual (Hedges, 1984, 1985; Bullock & Svyantek, 1985; Bangert-Drowns, 1986; Sánchez Meca y Ato, 1989; etc.), los que ayudan a los evaluadores a identificar variaciones que afectan a los resultados, creando nuevos análisis estadísticos que permiten una revisión cuantitativa, que, hoy por hoy, parece ser la mejor. La afirmación de Sánchez-Meca y Ato (1989) acerca de que la utilidad del meta-análisis se verá acrecentada en el futuro si las políticas editoriales y los investigadores primarios se conciencian de la importancia de la disponibilidad pública de los datos, la trasladamos al ámbito evaluativo referida también a dicha disponibilidad, que debería incluir el informe completo (con todos los anexos y datos) correspondiente a la evaluación del programa, y máxime si se ha seguido el modelo no lineal de evaluación.

La meta-evaluación tiene defensores y detractores muy radicales, e incluso se están proponiendo alternativas, como la de Slavin (1986), que añade a los sistemas tradicionales criterios de selección de los estudios a considerar, y el uso del tamaño de efecto (en lugar de la significación estadística solamente) como métrica común para los efectos de tratamiento.

Dado que en evaluación de programas se sigue un esquema de trabajo similar al de cualquier estudio empírico, aunque con las adaptaciones lógicas, hacemos extensivas a este ámbito las múltiples técnicas disponibles.

4 Evaluabilidad de un programa

Desde la perspectiva de Wholey (1987), la valoración de la evaluabilidad es un instrumento diagnóstico y prescriptivo que puede utilizarse para determinar la medida en que perviven algunos de los diferentes problemas que pueden plantearse, y especialmente los de:

- Falta de definición del programa;
- de lógica de las suposiciones que permiten comprobar los vínculos entre recursos, implementación, efectos e impacto;
- de acuerdo respecto a prioridades; y
- de incapacidad de actuación.

Si se halla presente alguno de los dos primeros, la evaluación será no concluyente o irrelevante; si se plantea el tercer o cuarto problema, aunque sea concluyente, la evaluación no podrá producir un incremento en el resultado o efectos.

El estudio de la evaluabilidad de un programa está muy relacionado con el de su teoría (Bickman, 1987), la cual identifica los recursos necesarios y disponibles para dicho programa, las actividades, el tipo de efectos que se esperan, y especifica la cadena de las supuestas relaciones causales planteadas. El estudio de la evaluabilidad de un programa implica un proceso (Schmidt, Scanlon & Bell, 1979; Rutman, 1980; Nay & Kay, 1982; Rog, 1985) mediante el cual se clarifica el punto de vista del autor y "gestionadores" respecto a la finalidad del programa, se explora la realidad del programa respecto a la plausibilidad de los objetivos y la viabilidad de la medida de los efectos, se analizan la documentación del programa y sus fuentes de información, y se estudian las posibilidades de cambio en caso de que haya lugar.

Como esquema básico, se han propuesto criterios de evaluabilidad (Shipman, 1989) que se presentan en un formato estandar que incide sobre los tres pilares fundamentales de necesidad del programa, su implementación y efectos alcanzados, aunque su excesiva generalidad dificulta una actuación objetiva y concreta. En este sentido, y de forma mucho más específica, Fernández Ballesteros (1988) ha elaborado un listado de cuestiones relevantes en evaluación de programas (LCREP) que tiene por objetivo el poder juzgar sobre la dificultad de la evaluación que va a ser abordada, con la ventaja de que cada una de las categorías de análisis se articula en una serie de variables, pudiéndose llegar a una cuantificación de la evaluabilidad en función del cumplimiento de una serie de criterios de calidad establecidos racionalmente.

5 Perspectivas futuras

Son muchas las cuestiones que quedan pendientes en evaluación de programas, referidas especialmente a la puesta a punto y adecuación de técnicas diversas, lo que es de prever que permitirá un desarrollo más completo y armónico. Hace unos años afirmaba Sperling (Wertheimer, Barclay, Cook, Kiesler, Koch, Riegel, Rorer, Senders, Smith & Sperling, 1978, p. 632) que "la frase mágica para el futuro será la evaluación de programas", en que se centrará especialmente la psicología de los años venideros, aunque para ello, claro está, será necesaria, en primer lugar, una previa tarea de elaboración de dichos programas, lo cual concuerda efectivamente con la tendencia creciente que venimos observando por parte de todo tipo de instituciones, y en segundo lugar, un avance en la valoración de su evaluabilidad (Cook, 1984).

Además existe un hecho indudable en el momento que estamos atravesando, y es que se están realizando serios progresos en evaluación de programas no como resultado de pequeñas aportaciones y sucesivos refinamientos, que también tienen su importancia, sino porque existe, en palabras de autores tan relevantes como Cordray y Lipset (1986, p. 17), "un cambio completo de paradigma", aunque ello implique un revulsivo que comporta controversia y polémica.

Controversia y polémica que en algunos casos se ha manifestado de forma virulenta, como en disciplina básica vs. aplicada, evaluadores internos vs. externos, o en uso de datos cualitativos vs. cuantitativos, o en estudios de casos vs. estudios experimentales, o en la construcción de determinados instrumentos, etc., pero que en la actualidad están ya cada vez más superados, adquiriendo con ello una madurez patente y manifiesta que le permitirá afrontar con garantías de éxito el reto que va a suponer la evaluación de programas en las próximas décadas.

6 Referencias bibliográficas

- Aldrich, W. & Nelson, S. (1984) *Linear probability, logit, and probit models*. Beverly Hills: Sage.
- Alvira, F. (1989) Diseños de investigación social: Criterios operativos. En García Ferrando, M., Ibáñez, J. y Alvira, F.(Eds.) *El análisis de la realidad social. Métodos y técnicas de investigación* (pp. 85-107). Madrid: Alianza.
- Anglin, M.D. & Fisher, D.G. (1987) Survival analysis in drug program evaluation. Part II. Partitioning treatment effects. *International Journal of Addictions*, 22 (4), 377-384.
- Anguera, M.T. (en prensa -a-) *Evaluación de programas desde una perspectiva metodológica*. Salud Mental Comunitaria.

- Anguera, M.T. (en prensa -b-) *Evaluación de programas de salud*. Salud Mental Comunitaria.
- Anguera, M.T. y Blanco, A. (1988-91) *Evaluación de programas en Servicios Sociales: Alternativas metodológicas*. Investigación subvencionada por la Comisión Interministerial de Ciencia y Tecnología (CICYT).
- Anguera, M.T., Blanco, A., Gil, A., Ibern, I., Pérez, E., Quera, V. y Redondo, S. (1989-90) *Instrumentos y técnicas de evaluación de programas de rehabilitación en el ámbito penitenciario*. Investigación subvencionada por la Dirección General de Servicios Penitenciarios y de Rehabilitación (Consejería de Justicia de la Generalidad de Cataluña).
- Anguera, M.T. y Redondo, S. (en prensa) La evaluación de la intervención penitenciaria. En S. Redondo: *Intervención en la prisión y su entorno. Un enfoque conductual aplicado*. Barcelona: Martínez Roca.
- Arnau, J., Anguera, M.T. y Gómez, J. (en prensa) *Metodología de investigación en Ciencias del Comportamiento*. Murcia: Servicio de Publicaciones de la Universidad.
- Artells, J.J. (1980, diciembre) L'avaluació econòmica de programes i polítiques sanitàries i serveis socials. *Acta de les Primeres Jornades sobre Economia de la Salut: La utilització del'anàlisi econòmica en els serveis sanitaris*, (pp. 251-287). Barcelona: Colegio de Economistas de Cataluña.
- Baltes, P.B. & Nesselroade, J.R. (1979) History and rationale of longitudinal research. In J.R. Nesselroade & P.B. Baltes (Eds.) *Longitudinal research in the study of behavior and development* (pp. 1-39). New York: Academic Press.
- Bangert-Drowns, R.L. (1986) Review of developments in meta-analytic method. *Psychological Bulletin*, 99, 388-399.
- Bauer, R.A. (Ed.) (1966) *Social indicators*. Cambridge, Mass.: M.I.T. Press.
- Bickman, L. (1987) The functions of program theory. In L. Bickman (Ed.) *Using program theory in evaluation* (pp. 5-18). San Francisco: Jossey-Bass.
- Boruch, R.F. & Pearson, R.W. (1988) Assessing the quality of longitudinal surveys. *Evaluation Review*, 12 (1), 3-58.
- Bullock, R.J. & Svyantek, D.J. (1985) Analyzing meta-analysis. Potential problems, an unsuccessful replication, and evaluation criteria. *Journal of Applied Psychology*, 70 (1), 108-115.
- Bunge, M. (1975) What is a quality of life indicator? *Social Indicators Research*, 2, 65-79.

- Burnam, M.A. & Koegel, P. (1988) Methodology for obtaining a representative sample of homeless persons. The Los Angeles Skid Row Study. *Evaluation Review*, 12 (2), 117-152.
- Burstein, L. & Freeman, H.E. (1985) Perspectives on data collection in evaluations. In L. Burstein, H.E. Freeman & P.H. Rossi (Eds.) *Collecting evaluation data* (pp. 15-32). Beverly Hills: Sage.
- Campbell, D.T. (1969) Reforms as experiments. *American Psychologist*, 24, 409-429.
- Campbell, D.T. (1982, March) Can we be scientific in applied social science?. Paper presented at the Annual Meeting of the American Educational Research Association. [Reprinted in R.F. Conner, D.G. Altman & C. Jackson (1984): *Evaluation Studies Review Annual*, vol. 9, pp. 26-48].
- Campbell, D.T. & Boruch, R.F. (1975) Making the case for randomized assignment to treatments by considering the alternatives. In C.A. Bennett & A.A. Lumsdaine (Eds.) *Central issues in social program evaluation* (pp. 195-297). New York: Academic Press.
- Campbell, D.T. y Stanley, J. (1967) *Diseños experimentales y cuasi experimentales en investigación educativa*. Buenos Aires: Amorrortu.
- Cannon, N.L., McGuire, T. & Dickey, B. (1985) Capital costs in economic program evaluation: The case of mental health services. In J.S. Catterall (Ed.) *Economic evaluation of public programs* (pp. 69-82). San Francisco: Jossey-Bass.
- Carley, M. (1983) *Social measurement and social indicators. Issues of policy and theory*. London: Allen & Unwin.
- Carmona, J.A. (1977) *Los indicadores sociales, hoy*. Madrid: Centro de Investigaciones Sociológicas.
- Carrasco, J.L. (1986) *El método estadístico en la investigación médica*. Madrid: Ciencia.
- Casas, F. (1988) *Indicadors psicosocials de risc a la infància*. Tesis Doctoral no publicada. Barcelona: Universidad de Barcelona.
- Cleave-Hogg, D. & Byrne, P.N. (1988) Evaluation of an innovation in a traditional medical school: A metaevaluation. *Evaluation & the Health Professions*, 11 (2), 249-271.
- Clifford, D.L. & Sherman, P. (1983) Internal evaluation: Integrating program evaluation and management. In A.L. Love (Ed.) *Developing effective internal evaluation* (pp. 23-45). San Francisco: Jossey-Bass.

- Conner, D.G. Altman & C. Jackson (Eds.) (1984) *Evaluation Studies Review Annual*, vol. 9 (pp. 26-48). Beverly Hills: Sage.
- Conner, R.F., Altman, D.G. & Jackson, C. (1984) 1984: A brave new world for evaluation? In R.F. Conner, D.G. Altman & C. Jackson (Eds.) *Evaluation Studies Review Annual*, vol. 9 (pp. 13-25). Beverly Hills: Sage.
- Conrad, K.J. & Miller, T.Q. (1987) Measuring and testing program philosophy. In L. Bickman (Ed.) *Using program theory in evaluation* (pp. 19-42). San Francisco: Jossey-Bass.
- Cook, T.D. (1981) Dilemmas in evaluation of social programs. In M.B. Brewer & B.E. Collins (Eds.) *Scientific inquiry and the social sciences. A volume in honor of Donald T. Campbell* (pp. 257-287). San Francisco: Jossey-Bass.
- Cook, T.D. (1984) Opportunities for evaluation in the next few years. *Evaluation News*, 5 (2), 20-46.
- Cook, T.D. & Campbell, D.T. (1979) *Quasi-Experimentation. Design and analysis issues for field settings*. Chicago: Rand McNally.
- Cordray, D.S., Bloom, H.S. & Light, R.J. (Eds.) (1987) *Evaluation Practice in Review*. San Francisco: Jossey-Bass.
- Cordray, D.S. & Lipsey, M.W. (1987) Evaluation Studies for 1986: Program evaluation and program research. In D.S. Cordray & M.W. Lipsey (Eds.) *Evaluation Studies Review Annual*, vol. 11 (pp. 17-44). Beverly Hills: Sage.
- Cordray, D.S. & Sonnefeld, L.J. (1985) Quantitative synthesis: An actuarial base for planning impact evaluations. In D.S. Cordray (Ed.) *Utilizing prior research in evaluation planning* (pp. 29-48). San Francisco: Jossey-Bass.
- Cox, D.R. & Oakes, D. (1984) *Analysis of survival data*. London: Chapman & Hall.
- Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D. Hornik, R.C., Phillips, D.C., Walker, D.F. & Weiner, S.S. (1980) *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Crowley, J. & Breslow, N. (1984) Statistical analysis of survival data. *Annual Review of Public Health*, 5, 385-411.
- Curry, R.H., Yarnold, P.R., Bryant, F.B., Martin, G.J. & Hughes, R.L. (1988) *Evaluation & the Health Professions*, 11 (1), 113-129.
- Chen, H.T., Quane, J., Garland, T.N. & Marcin, P. (1988) Evaluating an antismoking program: Diagnostics of underlying causal mechanisms. *Evaluation & the Health Professions*, 11 (4), 441-464.

- Davis, B.G. (Ed.) (1986) *Teaching of evaluation across the disciplines*. San Francisco: Jossey-Bass.
- Domínguez, H. (Dir.) (1985) *Curso sobre evaluaciones de impacto ambiental*. Madrid: M.O.P.U., Dirección General del Medio Ambiente.
- Duncan, G.J., Juster, F.T. & Morgan, J.N. (1984) The role of panel studies in a world of scarce research resources. In S. Sudman & M.A. Spaetg (Eds.) *The collection and analysis of economic and consumer behavior data: In memory of Robert Ferber*. Champaign, Il.: Bureau of Economic and Business Research.
- Eber, H.W. (1983) Multivariate methodologies for evaluation research. In E.L. Struening & M.B. Brewer (Eds.) *Handbook of evaluation research* (pp. 275-293). Beverly Hills: Sage.
- Fernández Ballesteros, R. (1987, septiembre) *Ciencia, ideología y política en evaluación de programas*. Conferencia pronunciada en el curso "Evaluación de programas y servicios para el bienestar social: Un enfoque pluridisciplinario". Santander: Universidad Internacional Menéndez Pelayo.
- Fernández Ballesteros, R. (1988) *Listado de cuestiones relevantes en evaluación de programas (LCREP)*. Manuscrito no publicado. Madrid.
- Fienberg, S.B. & Tanur, J. (1987) The design and analysis of longitudinal surveys: Controversies and issues of costs and continuity. In R.F. Boruch & R.W. Pearson (Eds.) *Designing research with scarce resources*. New York: Springer-Verlag.
- Finsterbusch, K. (1985) State of the art in social impact assessment. *Environment and Behavior*, 17 (2), 193-221.
- Finsterbusch, K., Llewellyn, L.G. & Wolf, C.P. (Eds.) (1983) *Social impact assessment methods*. Beverly Hills: Sage.
- Fisher, D.G. & Anglin, M.D. (1987) Survival analysis in drug program evaluation. Part I. Overall program effectiveness. *International Journal of Addictions*, 22 (2), 113-134.
- Fitzgerald, C.T. & Cutler, W. (1983) *Resource inventory and directory of services for Jackson County, Illinois*. Carbondale: Southern Illinois University, Applied Research Consultants.
- Forehand, G.A. (Ed.) (1982) *Applications of time series analysis to evaluation*. San Francisco: Jossey-Bass.
- Frederiksen, C.H. & Rotondo, J.A. (1979) Time-series models and the study of longitudinal change. In J.R. Nesselrode & P.B. Baltes (Eds.) *Longitudinal research in the study of behavior and development* (pp. 111-153). New York: Academic Press.

- Galarraga, J. (1980, diciembre) La demanda de cuidados médicos: Conceptos básicos y su cálculo empírico para Navarra. *Acta de les Primeres Jornades sobre Economia de la Salut: La utilització del'anàlisi econòmica en els serveis sanitaris* (pp. 47-64). Barcelona: Colegio de Economistas de Cataluña.
- Garfinkel, I. & Corbett, T. (1985) A design for an economic analysis: The Wisconsin Child Support Demonstration. In J.S. Catterall (Ed.) *Economic evaluation of public programs* (pp. 31-51). San Francisco: Jossey-Bass.
- Gold, N. (1981) *The stakeholder approach for educational program evaluation*. Washington: National Institute of Education.
- Goldstein, H. (1979) *The design and analysis of longitudinal studies. Their role in the measurement of change*. New York: Academic Press.
- Gottman, J.M. & Glass, G.V. (1978) Analysis of interrupted time-series experiments. In T.R. Kratochwill (Ed.) *Single subject research. Strategies for evaluating change* (pp. 197- 235). New York: Academic Press.
- Grabe, S. (1983) *Evaluation Manual*. Paris: Unesco.
- Gramlich, E.M. (1981) *Benefit-cost analysis of government programs*. Englewood Cliffs, N.J.: Prentice-Hall.
- Guba, E.G. & Lincoln, Y.S. (1981) *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey- Bass.
- Haggart, S.A. (1983) Determining the resource requirements and cost of evaluation. In M.C. Alkin & L.C. Solmon (Eds.) *The costs of evaluation* (pp. 59-70). Beverly Hills: Sage.
- Hedges, L.V. (1984) Advances in statistical methods for meta-analysis. In W.H. Yeaton & P.M. Wortman (Eds.) *Issues in data synthesis* (pp. 25-42). San Francisco: Jossey-Bass.
- Hedges, L.V. (1985) *Statistical methods for meta-analysis*. New York: Academic Press.
- Hsiao, C. (1986) *Analysis of panel data*. Cambridge: Cambridge University Press.
- Holland, P.W. & Rubin, D.B. (1988) Causal inference in retrospective studies. *Evaluation Review*, 12 (3), 203-231.
- Jöreskog, K.G. (1979) Statistical estimation of structural models in longitudinal-developmental investigations. In J.R. Nesselroade & P.B. Baltes (Eds.) *Longitudinal research in the study of behavior and development* (pp. 303-351). New York: Academic Press.

- Juan-Espinosa, M., Rubio, V. y Márquez, M.O. (1987) Inventario de recursos ambientales en el entorno de una institución de deficientes mentales. En R. Fernández Ballesteros (Coord.) *El ambiente. Análisis psicológico* (pp. 249-268). Madrid: Pirámide.
- Kaplan, E.H. & Abramson, P.R. (1989) So what if the program ain't perfect? A mathematical model of AIDS education. *Evaluation Review*, 13 (2), 107-122.
- Kessler, R.C. & Greenberg, D.F. (1981) *Linear panel analysis. Models of quantitative change*. New York: Academic Press.
- Kratochwill, T.R. (1978) Foundations of time-series research. In T.R. Kratochwill (Ed.) *Single subject research. Strategies for evaluating change* (pp. 1-100). New York: Academic Press.
- Landis, J.R. & Koch, G.G. (1979) The analysis of categorical data in longitudinal studies of behavioral development. In J.R. Nesselroade & P.B. Baltes (Eds.) *Longitudinal research in the study of behavior and development* (pp. 233-261). New York: Academic Press.
- Leistriz, F.L. & Murdock, S.H. (1981) *The socioeconomic impact of resource development. Methods for assessment*. Boulder, Co.: Westview Press.
- Levey, L.M., Nyman, J.A. & Haugaard, J. (1988) A benefit-cost cost analysis of family planning services in Iowa. *Evaluation & the Health Professions*, 11 (4), 403-424.
- Lewin, H.M. (1983) *Cost-effectiveness. A primer*. Beverly Hills: Sage.
- Lipsey, M.W., Crosse, S., Dunkle, J., Pollard, J. & Stobart, G. (1985) Evaluation: The state of the art and the sorry state of the science. In D.S. Cordray (Ed.) *Utilizing prior research in evaluation planning* (pp. 7-27). San Francisco: Jossey-Bass.
- Little, R.J. & Rubin, D.B. (1987) *Statistical analysis with missing data*. New York: Wiley.
- Mandell, M.B. (1989) Estimating the mediating effect of intervening variables in pooled cross-sectional and time series designs: Model specification and estimation procedures. *Evaluation Review*, 13 (2), 174-200.
- Martí-Costa, S. & Serrano-García, I. (1983) Needs assessment and community development: An ideological perspective. *Prevention in Human Services*, 3, 75-83.
- Maxim, P.S. (1989) The impact of design effects on standard errors in roadside traffic surveys. *Evaluation Review*, 13 (2), 157-173.
- Mayer, R.R. (1985) *Policy and program planning*. Englewood Cliffs, N.J.: Prentice-Hall.

- McCain, L.J. & McCleary, R. (1979) The statistical analysis of the simple interrupted time-series quasi-experiment. In T.D. Cook & D.T. Campbell. *Quasi-Experimentation. Design & Analysis Issues for Field Settings* (pp. 233-293). Chicago: Rand McNally.
- McCutcheon, A.L. (1987) *Latent class analysis*. Beverly Hills:Sage.
- McKillip, J. (1987) *Need analysis. Tools for the human services and education*. Beverly Hills: Sage.
- Mishan, E.J. (1988) *Cost-benefit analysis*. London: Unwin.
- Nagel, S. (1983a) Nonmonetary variables in benefit-cost analysis. *Evaluation Review*, 7, 37-64.
- Nagel, S. (1983b) Unknown variables in policy/program evaluation. *Evaluation and Program Planning*, 6, 7-18.
- Nagel, S. (1985) Economic transformations of nonmonetary benefits in program evaluation. In J.S. Catterall (Ed.) *Economic evaluation of public programs* (pp. 53-68). San Francisco: Jossey-Bass.
- Nay, J. & Kay, P. (1982) *Government oversight and evaluability assessment*. Lexington, Mass.: Heath.
- Neigher, W.D. & Metlay, W. (1983) Values and methods: Evaluation and management perspectives. In A.L. Love (Ed.) *Developing effective internal evaluation* (pp. 47-60). San Francisco: Jossey-Bass.
- Neuber, K.A. (1980) *Needs assessment. A model for community planning*. Beverly Hills: Sage.
- Olsen, M. & Merwin, D.J. (1977) Toward a methodology for conducting social impact assessment using quality of social life indicators. In K. Finsterbusch & C.P. Wolf (Eds.) *Methodology of social impact assessment*. Stroudsburg, Penn.: Dowden, Hutchinson & Ross.
- Palmer, A. (1988) *Análisis de la supervivencia*. Barcelona: Documentos del Laboratorio de Psicología Matemática de la Universidad Autónoma.
- Palumbo, D.J. (1987) Politics and evaluation. In D.J. Palumbo (Ed.) *The politics of program evaluation* (pp. 12-46). Beverly Hills: Sage.
- Pancer, S.M. & Westhues, A. (1989) A developmental stage approach to program planning and evaluation. *Evaluation Review*, 13 (1), 56-77.
- Patton, M.Q. (1978) *Utilization-focused evaluation*. Beverly Hills: Sage.

- Petersen, V.J., Bosanac, E.M., Baranowski, T. & Forren, G.L. (1981) *A resource inventory approach to needs assessment: Examples from a statewide hypertension control program*. Morgantown, WV: Department of Community Medicine, Office of Health Service Research.
- Pitz, G.F. & McKillip, J. (1984) *Decision analysis for program evaluators*. Beverly Hills: Sage.
- Plewis, I. (1985) *Analysing change. Measurement and explanation using longitudinal data*. New York: Wiley.
- Pollard, W.E. (1986) *Bayesian statistics for evaluation research*. Beverly Hills: Sage.
- Posavac, E.J. & Carey, R.G. (1985) *Program evaluation. Methods and case studies*. Englewood Cliffs, N.J.: Prentice-Hall.
- Reiss, A.J. (1985) Some failures in designing data collection that distort results. In L. Burstein, H.E. Freeman, & P.H. Rossi (Eds.) *Collecting evaluation data. Problems and solutions* (pp.161-177). Beverly Hills: Sage.
- Riecken, H.W. & Boruch, R.F. (Eds.) (1974) *Social experimentation. A method for planning and evaluating social intervention*. New York: Academic Press.
- Rodríguez Castedo, A. (1987, septiembre) *La evaluación en Servicios Sociales*. Conferencia pronunciada en el curso "Evaluación de programas y servicios para el bienestar social: Un enfoque pluridisciplinario". Santander: Universidad Internacional Menéndez Pelayo.
- Rog, D.J. (1985) *A methodological analysis of evaluability assessment*. Unpublished Doctoral dissertation. Vanderbilt University.
- Rogosa, D. (1979) Causal models in longitudinal research: Rationale, formulation, and interpretation. In J.R. Nesselroade & P.B. Baltes (Eds.) *Longitudinal research in the study of behavior and development* (pp. 263-302). New York: Academic Press.
- Rossi, P.H. (Ed.) (1982) *Standards for evaluation practice*. San Francisco: Jossey-Bass.
- Rossi, P.H. & Freeman, H.E. (1986) *Evaluation. A systematic approach*. Beverly Hills: Sage.
- Rossi, P.H. & Wright, J.D. (1987) Evaluation research. An assessment. In D.S. Cordray & M.W. Lipsey (Eds.) *Evaluation Studies Review Annual, vol. 11* (pp. 48-69). Beverly Hills: Sage.
- Rubin, D.B. (1974) Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology, 66*, 688-701.

- Rubin, D.B. (1987) *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rubio, V. (1987) Evaluación del impacto ambiental. En R. Fernández Ballesteros, R. (Coord.) *El ambiente. Análisis psicológico* (pp. 184-199). Madrid: Pirámide.
- Rutman, L. (1980) *Planning useful evaluations: Evaluability assessment*. Beverly Hills: Sage.
- Sackett, G.P. (Ed.) (1978) *Observing Behavior: Data collection and analysis methods*. Vol II. Baltimore: University of Park Press.
- Sackett, G.P. (1980) Lag sequential analysis as a data reduction technique on social interaction research. In D.B. Sawin, R.C. Hawkins, L.O. Walker & J.H. Penticuff (Eds.) *Exceptional infant. Psychosocial risks in infant-environment transactions* (pp. 300-340). New York: Brunner/Mazel.
- Sackett, G.P. (1987) Analysis of sequential social interaction data: Some issues, recent developments, and a causal inference model. In J.D. Osofsky (Ed.) *Handbook of infant development* (pp.855-878). New York: Wiley.
- Sánchez-Meca, J. y Ato, M. (1989) Meta-análisis: Una alternativa metodológica a las revisiones tradicionales de la investigación. En J. Arnau y H. Carpintero (Coords.) *Tratado de Psicología General. Vol. 1: Historia, teoría y método* (pp. 617-669). Madrid: Alhambra.
- Saturno, P.J. y Marset, P. (1980, diciembre) Necesidad y demanda de hospitalización: Influencia de la localización de recursos y el tamaño poblacional en la utilización de los servicios hospitalarios. Estudio de la distribución espacial de la frecuentación hospitalaria en 32 municipios de Murcia. *Acta de les Primeres Jornades sobre Economia de la Salut: La utilització de l'anàlisi econòmica en els serveis sanitaris* (pp. 91-107). Barcelona: Colegio de Economistas de Cataluña.
- Schaloch, R.L. & Thornton, C.V. (1988) *Program evaluation. A field guide for administrators*. New York: Plenum Press.
- Scheldon, E. & Freeman, H.E. (1970) Notes on social indicators: Promises and potential. *Policy Sciences, 1*, 97-111.
- Schmidt, R.E., Scanlon, J.W. & Bell, J.B. (1979) *Evaluability assessment: Making public programs work better*. Rockville, Md.: Project Share, Department of Health and Human Services.
- Scriven, M. (1973) Goal-free evaluation. In E.R. House (Ed.) *School evaluation: The politics and process*. Berkeley: McCutchan.
- Shipman, S. (1989) General criteria for evaluating social programs. *Evaluation Practice, 10 (1)*, 20-26.

- Slavin, R.E. (1986) Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15, 5-11.
- Sohm, E.D. (1978) *Glossary of evaluation terms*. Geneva: Joint Inspection Unit.
- Stake, R.E. (Ed.) (1975) *Evaluating the arts in education: A responsive approach*. Columbus: Merrill.
- Stokey, E. & Zeckhauser, R. (1978) *A primer for policy analysis*. New York: Norton.
- Stufflebeam, D.L. (1978) Meta-evaluation: An overview. *Evaluation & the Health Professions*, 1 (1), whole number.
- Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, L.R., Merriman, H.O. & Provus, M.M. (1971) *Educational evaluation and decision-making*. Itasca: Peacock.
- Sullivan, J.L. & Feldman, S. (1979) *Multiple indicators. An introduction*. Beverly Hills: Sage.
- Trochim, W.M. (1984) *Research design for program evaluation. The regression-discontinuity approach*. Beverly Hills: Sage.
- Veney, J.E. & Kaluzny, A.D. (1984) *Evaluation and decision making for health services program*. Englewood Cliffs, N.J.: Prentice-Hall.
- Wampold, B.E. & Holloway, E.L. (1983) A note on interobserver reliability for sequential data. *Journal of Behavioral Assessment*, 5 (3), 217-225.
- 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. & Sperling, S.E. (1978) Psychology and the future. *American Psychologist*, 33 (7), 631-647.
- Wholey, J.S. (1987) Evaluability assessment: Developing program theory. In L. Bickman (Ed.) *Using program theory in evaluation* (pp. 77-92). San Francisco: Jossey-Bass.
- Williams, W. (1975) Implementation analysis and assessment. *Policy Analysis*, 1 (3), 531-566.
- Wolf, R.L. (1979) The use of judicial evaluation methods in the formulation of educational policy. *Educational Evaluation and Policy Analysis*, 1, 19-20.
- York, R.O. (1982) *Human service planning: Concepts, tools and methods*. Chapel Hill: University of North Carolina Press.