

# Modelos teóricos para la práctica de la evaluación de programas

---

Xavier Ballart

---

Publicado en:

Q. Brugué y J. Subirats (eds.) *Lecturas de gestión pública*,  
Instituto Nacional de Administración Pública, Madrid, 1996  
Páginas 321-352; ISBN: 84-340-0907-2

# Modelos teóricos para la práctica de la evaluación de programas

Xavier Ballart

Universidad Autónoma de Barcelona

## I. Introducción

¿Qué valor público añade a la sociedad la promoción pública de viviendas, la construcción de nuevas infraestructuras de transporte rápido o la inversión en protección medio ambiental? ¿Qué se obtiene de la inversión en I+D, de la subvención de los estudios universitarios o de un programa de guarderías para madres trabajadoras? ¿Qué cabe esperar de los servicios preventivos de seguridad, sanitarios o sociales?

La evaluación de los programas públicos - y de las políticas en que estos se fundamentan - es una disciplina científica que se justifica por la presión que experimentan las instituciones públicas por determinar en qué medida la intervención pública produce una mejora en el bienestar individual o social, cómo se produce esta mejora y cómo se podría conseguir de una forma mas efectiva.

A diferencia del sector privado donde existe un criterio fundamental - el criterio del beneficio - para valorar la actividad de las personas y de las organizaciones, en el sector público, raramente existen unos criterios claros y ampliamente aceptados sobre los que fundamentar la valoración de la actividad de las instituciones político-administrativas. Intuitivamente podemos tener una cierta idea de si se produce igualdad en el acceso a determinado bien público, si una solución es económicamente eficiente, o sobre el nivel de pobreza en un determinado caso en relación con otros, pero resulta muy difícil comparar, en terminos de bienestar general, distintas “cantidades” de estos valores. La monetarización de este tipo de valores resulta muy difícil, lo que complica enormemente el análisis. La adopción de un criterio en perjuicio de otro se debe justificar, es preciso elaborar medidas del grado en que se consiguen estos objetivos, además de encontrar la manera en que se pueden introducir mejoras.

En la literatura sobre evaluación se señala su carácter aplicado y práctico en la medida que el objeto de estudio está determinado por los programas y las políticas de las instituciones públicas. Asimismo se señala el carácter de “arte”, ante las limitaciones que encuentra una aproximación “científica” al producirse la investigación en un contexto real. Sin embargo, ello no significa que la evaluación de programas carezca de una base teórica que se ha ido desarrollando a partir de las teorías básicas de otras disciplinas y de la acumulación de experiencias en áreas sectoriales específicas. Sería un error infravalorar la importancia de la teoría en la evaluación de programas y desaprovechar las lecciones de la experiencia.

Este artículo se divide en dos partes. La primera parte es una síntesis de las aportaciones de siete autores fundamentales para el desarrollo teórico de la evaluación de programas sobre la base de la selección de Shadish, Cook y Leviton (1991) - en el mismo sentido, Shadish y Epstein en un trabajo previo (1987) -. Esta selección de autores permite resumir en pocas

páginas la evolución histórica de la disciplina y los cambios que se han ido produciendo en el campo conceptual, introduciendo a su vez referencias a los trabajos de otros autores que elaboran más extensamente algunas de las ideas propuestas inicialmente por los autores seleccionados. En la segunda parte de este artículo, se elabora una categorización de las distintas dimensiones teóricas y prácticas que aparecen en las obras de estos autores con el objeto de explicar las tendencias en el desarrollo de la teoría de la evaluación y ofrecer un marco útil para encuadrar su ejercicio por académicos y/o consultores profesionales.

## II. Los fundamentos teóricos

La evaluación de programas es una disciplina cuyo origen y desarrollo inicial se produce en Estados Unidos, fundamentalmente a raíz de la legislación federal que obliga, y financia a la vez, la evaluación de los programas que impulsa. Esta es la base sobre la que se establece la profesión, lo que explica la variedad de enfoques y perspectivas. Los primeros programas cuya evaluación fue ordenada por la legislación federal fueron en educación si bien rápidamente este requisito se extendió a otros sectores: formación ocupacional, lucha contra la pobreza, delincuencia juvenil, vivienda.

Los siete autores seleccionados proceden de tradiciones científicas diversas y han desarrollado su profesión en departamentos universitarios, institutos o administraciones dedicados a distintas áreas: Campbell, Stake y Cronbach proceden de la psicología. Sin embargo, Stake al igual que Weiss, ha desarrollado su carrera académica en departamentos de educación. En el caso de Weiss, socióloga de formación, su vinculación como investigadora o consultora con distintos programas de la administración federal norteamericana, le dan una visión más amplia no circunscrita a temas de educación. Lo mismo cabe afirmar de Wholey, un matemático cuya carrera profesional se desarrolla fundamentalmente en la administración federal. Por último, Rossi es uno de los sociólogos más prestigiosos de los Estados Unidos mientras que Scriven es un filósofo de la ciencia.<sup>1</sup>

### 1. Scriven. La ciencia de la valoración.

Scriven es uno de los primeros teóricos de la evaluación. Su artículo “The methodology of evaluation” (1967,1972) es uno de los artículos más citados en evaluación educativa (Smith,1981 cit. por Shadish, Cook y Leviton, 1991). En este artículo introduce los términos evaluación formativa y evaluación sumativa, para distinguir entre las evaluaciones cuya realización tiene por objetivo proporcionar información a las personas que intentan mejorar el objeto evaluado, de las evaluaciones pensadas para proporcionar información sobre si se debe continuar o terminar un programa. Scriven muestra una clara preferencia

---

<sup>1</sup> Ante la dificultad de hacer una selección, he preferido seguir el criterio de Shadish, Cook y Leviton. La falta de autores europeos no debería sorprender al lector. Las aportaciones teóricas originales en evaluación se deben fundamentalmente a autores norteamericanos. Los trabajos europeos en evaluación son fundamentalmente obras de divulgación o de aplicación a casos concretos. En España, los trabajos de Alvira (1992) o los míos propios (Ballart 1992, 1994; Ballart y Riba, 1996) entrarían dentro de estas categorías. Algunos artículos recientes de revisión de la evaluación en otros países europeos (Pollit, 1992, para Inglaterra; Duran, Monnier, Smith, 1995, para Francia) confirman esta apreciación.

por la evaluación sumativa en la medida que la primera requiere un conocimiento que va mas allá de si un programa ha funcionado mejor o peor, lo que no siempre es factible.

Scriven es el autor que más ha profundizado en la idea de evaluación como ciencia de la valoración. Critica a los que definen evaluación como producir información para la toma de decisiones. “evaluación es lo que es, la determinación de mérito o valor, para lo que sea utilizada es otra cuestión” (1980, p.7). “La investigación evaluativa debe producir como conclusión exactamente el tipo de afirmación que a los científicos sociales se les ha dicho es ilegítima: un juicio de valor o mérito” (1974, p.4).

Scriven entiende que la sociedad reclama una ciencia de la valoración porque la sociedad necesita saber si su administración, sus funcionarios y sus programas son buenos. Un programa será un buen programa si satisface las necesidades más importantes de la población. “La evaluación debería permitir determinar la necesidad que atiende el rendimiento actual del programa y los rendimientos que mejor darían satisfacción a grupos de necesidades” (1981, p.130).

Para Scriven, la evaluación debería servir a los intereses de los afectados por el programa que está siendo evaluado. La evaluación no debe hacerse desde la perspectiva de la gestión del programa, ni tampoco desde la perspectiva del usuario o cliente. Se trata de una perspectiva que debería estar por encima de estos dos puntos de vista, algo que podríamos llamar “el interés general” y que sin embargo señala para cada una de las partes la importancia del punto de vista de las demás (1980, p.103)

La evaluación de un programa no siempre está justificada: las evaluaciones deberían satisfacer el criterio de coste-eficacia: los beneficios fiscales de hacer una evaluación deben exceder los costes de su realización, en caso contrario no debería hacerse (1976 , p.220).

Scriven quiere minimizar la posibilidad de sesgo al formular un juicio de valor. Para ello identifica algunas razones por las que se tiende a sesgar la evaluación: el contexto organizativo puede llevar al evaluador a poner los intereses de la administración por encima de los demás. La lealtad al programa o a la institución que lo financia, la pertenencia como personal a la organización (evaluación interna), o la co-optación del evaluador por el programa convirtiéndolo en un defensor del mismo, son algunos de los elementos contextuales que pueden estar en el origen de este problema.

La tendencia a evaluar los programas en función de los objetivos de sus responsables y técnicos es la otra principal causa de sesgo en la valoración, según Scriven. Para este autor, la evaluación debe servir para identificar todos los efectos, intencionados o no, que puedan ayudar a resolver los problemas sociales. Los grandes objetivos de un programa normalmente son vagos y abstractos dado que lo que pretenden es obtener apoyo político, mientras que los objetivos que se fijan profesionales y técnicos para un programa también suelen estar sesgados dado que se seleccionan aquellos que es más factible conseguir. Por todo ello, Scriven es un claro partidario de la evaluación goal free, en función de las necesidades y no de los objetivos afirmados, Scriven recuerda que la justicia es ciega al igual que la medicina, que lo es doblemente. En evaluación, ni el paciente, ni el medico, ni

el evaluador conocen de antemano que efectos cabe esperar. “El evaluador tiene que descubrir los efectos” de un programa y comparar “los efectos con las necesidades de aquellos a los que afecta” (1983, p. 235).

Otras formas de controlar los posibles sesgos es haciendo circular los resultados de una evaluación a todas las partes afectadas, dando tiempo para su discusión. Esta ya es una forma de metaevaluación, otro término acuñado por Scriven para referirse a la necesidad de que las evaluaciones sean a su vez evaluadas. Para ello Scriven utiliza una lista de cuestiones (objeto, cliente, contexto, recursos disponibles, función, beneficiarios, necesidades y valores, estándares, limitaciones, resultados, posibilidad de generalizar, coste...) que es preciso tener en cuenta antes de iniciar una evaluación (1980, pp. 113-116).

Para Scriven es importante dedicar atención a la lógica de la evaluación para mejorar su práctica profesional. Para este autor, evaluar programas implica tres actividades cruciales:

1. determinar criterios, esto es, las dimensiones en las que el programa es evaluado, para lo que recomienda comprender bien su naturaleza: “entender lo que es un reloj permite automáticamente entender los criterios o dimensiones adecuados para juzgar su valor o mérito: exactitud, legibilidad, solidez...”
2. determinar los niveles de rendimiento aceptables en cada una de las dimensiones, utilizando estándares absolutos o comparativos; y
3. medir el rendimiento para el caso concreto, comparando los niveles alcanzados con los de referencia. Para determinar si vale la pena hacer el análisis habrá que considerar “el tiempo y recursos necesarios para la investigación, la probabilidad de obtener alguna respuesta y la probabilidad de que la respuesta sea correcta (1980, p.45)

Todo ello lleva a elaborar un juicio de valor sobre el objeto evaluado. Para Scriven, la cuestión fundamental es determinar si se trata de un buen programa, si vale la pena mantenerlo, qué aspectos funcionan adecuadamente y que es lo que resuelven. La mejor manera de obtener respuestas a cuestiones sobre efectos la proporcionan los métodos experimentales. Sin embargo no es un gran defensor de éstos en la medida que, a menudo, la adherencia al método lleva a responder preguntas irrelevantes. En estos casos prefiere una respuesta menos refinada pero a una cuestión mas relevante.

## 2. Campbell. La sociedad experimental.

La obra de Campbell tiene un ámbito de aplicación que va mas allá de la evaluación de programas. Sin embargo, su trabajo junto con Stanley, *Experimental and Quasi-Experimental Designs for Research* (1963) ha sido considerado como el que mayor influencia ha tenido en el desarrollo del campo (Shadish y Epstein, 1987). En él, Campbell y Stanley presentan una teoría de la causación, que se estructura en base a los tipos de validez y a las posibles amenazas a la validez de una investigación y definen los diseños cuasi experimentales como un tipo de investigación mas factible al poder llevarse a cabo en contextos reales y no depender de la normalmente difícil asignación aleatoria de los tratamientos. De hecho, Campbell y Stanley analizan las funciones que cumple la

asignación aleatoria en los verdaderos experimentos para explorar como se pueden conseguir de otra manera.

La teoría de la causación de Campbell se diferencia de las llamadas teorías esencialistas en que no pretende predecir de forma absoluta una variable dependiente (los resultados de un programa) sino simplemente estimar el impacto marginal de una variable independiente (programa) en aquella, cuando otras posibles causas se mantienen constantes a través del proceso de aleatorización. Campbell y Stanley adoptan el término validez interna para referirse a las inferencias que puedan relacionarse con agentes causales manipulables y explican nueve amenazas a la validez interna, dedicando una especial atención a la selección y a sus subcategorías - maduración e historia -.

Su principal contribución fue, sin embargo, la descripción y evaluación de algunos diseños sin aleatorización - cuasi experimentos - pero que sin embargo facilitan la inferencia causal. Estos se fundamentan en dos estrategias: primero, la utilización de medidas pretest (cuanto mas larga sea la serie de observaciones pretest, mejor), y segundo, la utilización de grupos de comparación lo más similares que se puedan formar como base de referencia para el no tratamiento. De la combinación de estas dos estrategias resultan múltiples posibilidades de diseños de investigación aplicables en la práctica.

Campbell prioriza la validez interna por encima de la validez externa. ¿De qué nos sirve generalizar una relación si dudamos de la existencia de la misma? Pero Campbell es quien adopta el término validez externa y explica siete amenazas que pueden circunscribir una relación de causalidad a las características de un grupo, a una localización, momento o forma de desarrollar un programa. Campbell y Stanley advierten, sin embargo, sobre las limitaciones de los cuasi experimentos y dejan bien claro que la mejor opción es para ellos la del experimento aleatorio.

Una buena parte del trabajo posterior de Campbell con otros investigadores y singularmente con Cook (1979) consiste en el análisis de la aplicación de los cuasi experimentos. En un trabajo con Boruch (1975), advierte sobre las limitaciones del diseño cuasi experimental más utilizado - el diseño de pretest-post test con grupos de comparación no equivalentes - y presenta una lista de experimentos desarrollados en ciencias sociales con el objeto de demostrar su factibilidad. Campbell defiende los experimentos argumentando que las dificultades que presentan en términos de coste, complejidad técnica y desarrollo temporal lento son superables.

Para Campbell, la administración debería evaluar una nueva política sometiéndola a un experimento antes de hacerla extensiva a toda la población. Ello permitiría someter a prueba los programas claramente innovadores - dada la escasa relevancia política de una prueba piloto - además de permitir establecer inferencias causales más claras.

El optimismo inicial de Campbell disminuiría a raíz de algunas críticas al método experimental en el contexto político-administrativo y ante la falta de voluntad política para llevar a cabo verdaderas reformas en el campo social y dar contenido, más allá de la retórica política, a la obligación de evaluar las políticas impulsadas por la legislación federal (1982, p.118). En *Reforms as Experiments* (1969) defiende aquellos cuasi

experimentos que mejor pueden informar sobre relaciones causales. Concretamente, destaca el diseño de series temporales interrumpidas donde la intervención pública puede cambiar la tendencia de la serie y el diseño de discontinuidad en la regresión donde se asigna un determinado tratamiento sobre la base de puntos de corte en una variable cuantitativa.

*Reforms as Experiments* fue criticada por Weiss y Rein (1970) quienes encontraron dificultades en la implementación de la evaluación experimental de un proyecto piloto. Sobre la base de observaciones cualitativas, Weiss y Rein critican las evaluaciones que no informan sobre lo que ocurre en la “caja negra”, refiriéndose con este término al proceso de implementación del programa. Campbell (1970), sin embargo, tiene una posición post positivista y rechaza la idea de una teoría neutral o de un conocimiento objetivo. Al contrario defiende que la medida cuantitativa se basa en suposiciones cualitativas: Distintos actores prefieren distintos resultados y ello tiene consecuencias diversas en términos de conceptualización teórica. Por ello, rechaza la investigación que se fundamenta en una sola medida de resultados y, al contrario, propone la utilización de múltiples indicadores y el análisis de los procesos para superar el problema de la caja negra. Campbell no discute pues la oportunidad de los análisis cualitativos, pero si se pronuncia en contra de que éstos substituyan a los métodos cuantitativos en la función de análisis causal.

Después de *Reforms as Experiments*, Campbell escribió sobre la llamada “Sociedad Experimental”, un artículo cuya publicación definitiva retrasaría unos años, en el que defiende su visión de una sociedad racional en la que las decisiones se toman después de evaluar distintas alternativas en la forma de hacer frente a los problemas sociales. Campbell cree que la formulación de las alternativas depende de los grupos políticos. Es el proceso político el que ha de asignar pesos a los indicadores y sacar conclusiones. El papel del científico social no debe ser pues el de consejero sino el de metodólogo. “The Experimenting Society” (1988), que es como se acabaría titulado el artículo, se fundamenta en los principios de la multiplicidad de análisis, crítica abierta y réplica de los estudios siempre que sea posible.

### 3. Weiss. El contexto político y organizativo. La utilización de los resultados de la investigación social.

Weiss es una socióloga que adopta las teorías y métodos experimentales en su primera obra pero que pronto advierte sobre las dificultades políticas y organizativas que presenta la evaluación para producir resultados que sean efectivamente utilizados. Esta aproximación es desarrollada, entre otros, por Chelimsky (1987) y Palumbo (1987).

Weiss describe el pesimismo y la decepción que producen los resultados de la primera oleada de evaluaciones en los años 70. “Nada parecía funcionar como se esperaba...Ocasionalmente se detectaban efectos positivos. Algunos programas mostraban pequeños beneficios...Pero incluso las evaluaciones más recientes, utilizando criterios racionales y métodos de estudio más sofisticados tienden a encontrar éxitos marginales” (1987, p.41). Por otra parte, “los datos recogidos parecían no tener ningún efecto en las

decisiones presupuestarias o sobre la expansión o reducción de los programas” (1987, p.42).

La teoría de Weiss responde a esta preocupación tomando como hilo conductor el problema de la insuficiente utilización de los resultados de la evaluación y de la investigación social en el proceso de formulación de las políticas. “En su forma ideal, la evaluación se lleva a cabo para un cliente que tiene que tomar decisiones y que mira a la evaluación en busca de respuestas en las que fundamentar sus decisiones,” (1972b, p.6).

Para Weiss, existe una incompatibilidad básica entre investigación social y programación social que explica los pobres resultados en términos de utilización. Su argumentación se fundamenta en las siguientes ideas:

1. La investigación no es el primer objetivo de los programas sociales. Cuando las necesidades de la evaluación y las necesidades del programa entran en conflicto, la prioridad es para el programa.
2. Los objetivos de los programas tienden a ser generales, difusos, diversos e inconsistentes. Varían para cada una de las partes afectadas y tienen poco que ver con el funcionamiento del programa. “A menudo, es preciso obtener el apoyo de una coalición para garantizar la aprobación de un programa. Hay que ganar la confianza de actores con valores e intereses distintos lo que lleva a hacer promesas realistas y no tan realistas” (Weiss, 1973a p.181).
3. Los técnicos y el personal de un programa suelen mostrar resistencia ante una evaluación. Su principal objetivo es producir el servicio mientras que una evaluación puede ser un obstáculo a sus tareas. Pueden pensar que la evaluación no es necesaria o que se les está juzgando a ellos, con lo que aún colaborarán menos.
4. Los grupos de control son factibles en laboratorios pero son más difíciles de organizar en un contexto real. La administración puede rechazar la asignación aleatoria de tratamientos por razones éticas o de otro tipo, puede que todas las personas que precisen el tratamiento lo reciban, y, aun en el caso de que no haya suficientes recursos para todos, cuando se produce una baja, la administración tiene que aceptar a una persona que estaba en el grupo de control.
5. En un laboratorio se puede homogeneizar el tratamiento que recibe cada individuo. En un contexto real, una prestación puede variar significativamente de un caso a otro, de un lugar a otro, “incluso en el caso de que el impacto medio no sea significativo, algunos lugares, algunas estrategias o determinadas formas de operar pueden funcionar” (1973a, p.182).
6. Muchos programas cambian de forma y de contenido a lo largo de su existencia, con nuevos directores, nuevas reglamentaciones o cambios en la ideología dominante. Algunos programas son simplemente un “conglomerado de actividades cambiantes que requieren un gran esfuerzo de especificar y describir” (1972a, p.9). Si resulta difícil



describir un programa aún lo es más explicar qué componentes son responsables de determinados efectos.

7. En un laboratorio se puede esperar a que un tratamiento acabe, o por el contrario, si algo va mal, se puede cambiar. Un programa real, raramente se puede cambiar completamente. Las innovaciones deben ser consistentes con la ideología y con los intereses del contexto político-administrativo. Por ello, los cambios serán “incrementales antes que masivos” (1980, p.25).

Weiss advierte asimismo sobre el tiempo que a menudo debe transcurrir para que se puedan detectar los efectos de una intervención sobre un problema social. Por ello, estima poco realistas algunas evaluaciones que esperaban detectar resultados de forma inmediata. Aun otras veces, se suponía que el programa tendría un impacto tan fuerte que cambiaría la vida de las personas por encima de la influencia de otros factores.

Todo ello justificaría un cierto abandono de los métodos experimentales así como la adopción de métodos cualitativos. Sin embargo, estos también presentan sus limitaciones. El principal inconveniente reside, para Weiss, en la falta de credibilidad de estas aproximaciones. Sin evidencia empírica dura, los resultados de una evaluación serán interpretados como las “opiniones de un observador” (1987, p.43). Weiss recomienda utilizar métodos cuantitativos y cualitativos. Estos últimos pueden aportar la flexibilidad, dinamismo y variedad de perspectivas de la que carecen los primeros.

La teoría sobre la utilización de los resultados de las evaluaciones presta una gran atención al contexto político de los programas sociales. “La evaluación es una empresa racional que se desarrolla en un contexto político” (1973b, p.37). Las consideraciones políticas tienen entrada de tres formas distintas:

1. Los programas son el resultado de decisiones políticas... Son objeto de las presiones, a favor y en contra, que surgen a raíz del proceso político.
2. Los resultados de una evaluación entran en la arena política en la que compiten por la atención de los decisores con otros factores que tienen su peso en el proceso político.
3. La evaluación, en si misma, es política en tanto que implícitamente adopta determinadas posiciones políticas sobre la legitimidad de los objetivos o de la estrategia del programa, el carácter discutible de algunas cuestiones y no de otras... “Parece que la investigación en evaluación tiene más posibilidades de afectar decisiones cuando el investigador acepta los valores, suposiciones y objetivos del decisor” (p.41).

Otro elemento clave del contexto en el que se desarrolla la evaluación de programas es la existencia de múltiples partes afectadas por un programa. Un *stakeholder* es definido como el miembro de un grupo que es afectado de forma clara por el programa - y que por tanto puede verse afectado por las conclusiones de una evaluación - o el miembro de un grupo que toma decisiones sobre el futuro de un programa (1983a, p.84).

Cada grupo tiene intereses distintos en el programa, se hace preguntas distintas sobre el programa y tiene distintos niveles de poder para utilizar los resultados de una evaluación. Muchas partes no tienen ningún interés en descubrir si el programa funciona. Esta diversidad de *stakeholders* y de intereses hace mas difícil la utilización de los resultados de las evaluaciones

“Algunos justifican el programa porque ‘hace algo’ en un área en la que las necesidades son obvias. Otros se declaran satisfechos si el programa evita que un joven cometa un delito o prepara a un alumno para ir a la universidad. La administración, el personal, la clientela que se forma entorno a un programa tienen un interés en su perpetuación y en la expansión de la intervención” (1972a, p.4).

Ante esta visión mas bien pesimista de las posibilidades de utilización de las evaluaciones, Weiss orienta su trabajo, que desarrolla en parte con Bucuvalas, a analizar cómo se formulan las políticas y qué papel juega en este proceso la investigación social en un sentido más amplio. Para Weiss y Bucuvalas, raramente se utilizan los resultados de un estudio para dar respuesta a una pregunta específica que luego se implementa. Mas bien se produce una absorción de conceptos y generalizaciones de muchos estudios a lo largo de períodos de tiempo largos de manera que se produce una integración de los resultados de las investigaciones, junto con otras informaciones, en la interpretación que hacen los líderes políticos y administrativos de los hechos reales. Se produce así una “sensibilización gradual a las perspectivas de la investigación social” (1980, p.263).

Weiss defiende frente a una utilización instrumental de la investigación una utilización conceptual. Su *enlightening model* parte de la idea que la investigación no resuelve problemas sino que proporciona evidencias que pueden ser utilizadas para encontrar soluciones. En cada caso concreto, sin embargo, el investigador deberá considerar las llamadas *tres I* (de *ideología*, *intereses* presentes e *información* existente), ver como interaccionan y determinar si tiene sentido realizar un nuevo estudio (1983 b).

Finalmente, Weiss distingue tres etapas en toda investigación sobre un programa:

1. Formulación de las cuestiones que deben ser investigadas. Su recomendación es la de leerlo todo, hablar con todo el personal, observar el programa y colaborar con todas las partes con el fin de alcanzar un acuerdo. En esta etapa, el investigador debe construir un modelo de los procesos que el programa pretende generar.
2. Realización del estudio. Weiss señala aquí la importancia de la calidad de los estudios para la posterior utilización de sus resultados. En su primera época hablaba más de experimentos y cuasi experimentos. Posteriormente, acepta cualquier método (sondeos, estudios de caso, observaciones...), mientras se utilicen adecuada y correctamente. Desde la perspectiva de la organización de la investigación, el elemento crítico es para Weiss obtener la colaboración de las personas que trabajan en el programa, recogiendo las recomendaciones que se encuentran en la literatura.

3. Discusión de las implicaciones de la investigación para las políticas analizadas, señalando aquí la importancia de las recomendaciones y de la difusión de los resultados de la investigación (1972b).

#### 4. Wholey. Evaluación para la gestión.

La mayor parte de la obra de Wholey se refiere a la evaluación como una respuesta a la necesidad de los directivos y gestores públicos de evaluar sus programas con el objetivo de encontrar maneras de gestionarlos mejor. Abramson y Bellavita (en Wholey, Abramson y Bellavita, 1986) y en general, la literatura sobre *performance management* y control de gestión (Brizius y M.D. Campbell 1991, Dilulio, (1994) siguen esta aproximación.

La prioridad para Wholey esta en la gestión de los programas. La evaluación debería servir para facilitar mejoras evidentes en la gestión, rendimiento y resultados de los programas de la administración (1983). El destinatario son los directivos públicos, en un sentido amplio, que incluye a gestores y a otros directivos como los responsables presupuestarios o los mismos legisladores.

Al igual que Weiss, Wholey justifica la evaluación en términos de su utilidad: “El esfuerzo por evaluar un programa sólo se justifica si resulta en productos que se utilizan” (Abramson y Wholey, 1981, p.42). Sin embargo, Horst, Nay, Scanlon y Wholey (1974) encontraron poca evidencia de que la evaluación de programas sirviera para mejorar los programas de la administración federal. Para Wholey, los problemas detectados en la gestión y evaluación de los programas son inherentes a la complejidad del contexto administrativo y a las diferencias en la implementación de un lugar a otro. “La evaluación de grandes programas es difícil por el coste de recoger datos y por la variación en las actividades y objetivos a nivel local” (Wholey, 1983, p.124).

A ello hay que añadir el hecho de que la mayor parte de los programas son administrados por administraciones estatales o locales. “Muchos programas federales son grandes sobres de dinero para invertir en una área de problema...Un programa puede tener suficiente entidad para ser descrito por la prensa, para generar presión a favor o en contra de su existencia o para ser adoptado por la administración federal, y, sin embargo, las acciones que engloba esta intervención pueden no estar suficientemente especificadas (Horst, et al. 1974, p.303).

El principal problema para Wholey es de gestión. “Si definimos gestión como la aplicación de recursos a tareas con el objeto de conseguir objetivos específicos”, no se puede gestionar sin objetivos claros, hipótesis plausibles y demostrables sobre los resultados que se persiguen, y sin la motivación, habilidad o autoridad para gestionar. La actividad de evaluación está estrechamente relacionada con la gestión en el modelo de gestión orientada a resultados de Wholey. Concretamente, las actividades de evaluación se desarrollan en seis niveles:

1. En el trabajo con los niveles político y administrativo y con otros grupos de interés para identificar objetivos y prioridades y diagnosticar las potencialidades de un programa. En este nivel se llegan a acuerdos sobre los objetivos del programa y sobre los indicadores que serán utilizados para valorar el rendimiento del programa y sus resultados.
2. En el desarrollo de los sistemas necesarios para analizar el rendimiento del programa en función de los objetivos e indicadores de rendimiento especificados.
3. En el trabajo con decisores y otros grupos de presión para fijar un nivel realista de mejora posible del rendimiento del programa.
4. En el desarrollo de sistemas que faciliten la utilización de la información generada con el objeto de incentivar y premiar dentro de la organización, las mejoras en el rendimiento y en los resultados del programa.
5. En la valoración de los resultados observados y en su documentación, señalando cómo se consiguen unos buenos resultados así como los factores que pueden ser un obstáculo a la consecución de mejores resultados.
6. En la comunicación del rendimiento del programa a los niveles políticos y a la opinión pública en general (1983, p.202).

No todos los programas requieren, sin embargo, todas estas actividades de evaluación. Un estudio de evaluación solamente debería encargarse cuando “la utilidad probable de la nueva información supera los costes de obtenerla” (1983, p.119). Wholey propone cuatro tipos de estudio que pueden utilizar los gestores para determinar progresivamente si precisan una evaluación mas intensa, amplia y compleja.

#### a. Análisis de evaluabilidad (*Evaluability assessment*)

El objetivo es responder a la cuestión de si una evaluación contribuirá a mejorar el rendimiento de un programa (1983, p.35). Los productos del análisis de evaluabilidad son dos: un modelo del programa que asocie recursos, actividades, resultados y relaciones causales y un conjunto de indicadores de rendimiento sobre los que hay un mínimo acuerdo (1983, p. 42).

Además de clarificar la “teoría” del programa, el análisis de evaluabilidad sirve para explorar la realidad del programa y para asistir a directivos y legisladores a determinar opciones de cambio y de mejora.

#### b. Evaluación rápida (*rapid feedback evaluation*)

Un paso intermedio entre el análisis de evaluabilidad y la evaluación intensiva es la evaluación rápida con un doble objetivo: primero, hacer una valoración rápida del

rendimiento del programa en términos de los objetivos e indicadores de rendimiento acordados; y segundo, proponer algunos diseños para una evaluación más global, válida y fiable (1983, p.119).

La principal idea aquí es reducir el tiempo de feedback para los directivos y responsables políticos. La evaluación rápida solamente se puede hacer sobre la base de datos existentes, investigaciones previas, auditorías o informaciones recogidas en entrevistas, sondeos o visitas.

Otros autores (Hendricks, 1981) desarrollan los llamados *Service delivery assessments*, con la idea de producir información sobre un servicio sin las limitaciones que impone desarrollar un paquete de objetivos e indicadores previamente. Se trata de la descripción de las operaciones desarrolladas a nivel local, de su comparación con niveles de rendimiento considerados normales, de recoger información sobre problemas detectados y sobre las “mejores prácticas observadas”, para hacer recomendaciones sobre posibles mejoras operativas (1983, p. 144).

#### c. Control de resultados (*Performance monitoring*)

Se corresponde con el nivel 2 de gestión orientada a resultados. Consiste fundamentalmente en el establecimiento de un sistema de control de resultados sobre la base de un acuerdo a nivel político y administrativo de los objetivos e indicadores que se consideran realistas y aceptables (1983, p.155).

Wholey puntualiza que se trata simplemente de fijar la idea de rendimiento o resultados y medir el avance hacia los objetivos acordados, dejando para otros la tarea de estimar la medida en que las actividades del programa han causado los resultados estimados (1983, p. 155). Wholey entiende que un sistema de control de resultados es útil aún sin la posibilidad de establecer inferencias causales válidas. Se trata de un “sistema de señales” que sirve para el ajuste y la mejora continua del programa. Por otra parte, Wholey estima que, junto con los estudios de evaluación cualitativos, el control de resultados suele ser la alternativa más factible (1983, p.155).

Con un sistema de estas características existe la posibilidad de comparar datos con un período previo, con el rendimiento esperado si se han desarrollado estándares o entre distintas unidades de un mismo programa. Wholey advierte sobre la posible “corrupción” de los datos si se utilizan para tomar decisiones que pueden afectar negativamente a los que los proporcionan. También advierte sobre la necesidad de llegar a acuerdos sobre las actividades que serán objeto de control.

#### d. Evaluación intensiva

Evaluación intensiva es el estudio de evaluación más complejo posible. La principal preocupación es aquí la del diseño de la investigación (experimental o cuasi experimental) con el objeto de determinar la validez de las hipótesis causales que relacionan las actividades de un programa con sus resultados (1979, p. 49).

Sin embargo, es claro que para Wholey, el control de resultados a través de un sistema de indicadores es la opción más factible, menos cara y más útil para el gestor (1983, p.117). En el mismo sentido, una evaluación rápida puede ser la mejor opción antes de entrar en un proyecto de evaluación intensiva que puede no merecer el esfuerzo en tiempo y dinero.

Respecto a la organización de la evaluación, Wholey señala la oportunidad de integrar la oficina de evaluación con el programa, así como la importancia de aportar resultados pronto, mantener contactos frecuentes y, en general, compartir de forma constante, datos, resultados y conclusiones con el personal y los directivos del programa.

##### 5. Stake. Estudios de caso y métodos cualitativos.

Stake es uno de los primeros defensores de los métodos cualitativos en evaluación (1975a, 1975b, Stake y Gjerde 1974), y uno de los autores que más ha hecho por su legitimación. Guba y Lincoln (1981, 1989), Lincoln y Guba (1985), House, (1980), Patton (1980) siguen esta aproximación, desarrollando una de las líneas de investigación más seguidas en la práctica profesional. Al igual que Weiss y Wholey, Stake reacciona ante las primeras señales de no utilización de los resultados en evaluación. Su línea de argumentación pasa por defender la metodología de los estudios de caso como instrumento para mejorar la práctica de la evaluación a nivel local. Stake rechaza que las ciencias sociales puedan descubrir relaciones de causalidad input-output y no presta atención a los indicadores. Para Stake, los científicos sociales deben describir situaciones individuales que puedan ser comprendidas por los no científicos.

Al igual que Scriven, Stake deja que la evaluación emerja de la observación del programa “Desgraciadamente, solamente algunas cuestiones pueden ser estudiadas en una evaluación formal. El evaluador debe decidir en qué centra su atención, para ello, ¿debe basarse en sus percepciones?, ¿en los objetivos formales del programa?, ¿en las actividades del programa?, ¿en las reacciones de los participantes en el programa? La mayoría se basarán en una noción preconcebida de éxito. Mi recomendación es prestar atención a lo que ocurre en el programa, y después decidir las cuestiones de valor y los criterios (Stake, 1975b, p.15).

A diferencia de Wholey, Stake no orienta la evaluación en función de los intereses de los directivos encargados de la gestión. Su idea fundamental es la de identificar los intereses que distintas personas o grupos tienen en el programa y servir a los que el programa debería ayudar. También tiene una preocupación fundamental por ayudar al profesional que trabaja día tras día en el programa. El trabajo del evaluador no es pues, el de hacer un juicio sumativo. “Veo al evaluador en el papel de prestador de un servicio antes que en el rol de un filósofo-educador” (1975a, p.36).

Esta concepción de la evaluación como un servicio está en la base de su propuesta de un modelo de evaluación que responda a las necesidades de los clientes. Por *responsive evaluation*, entiende aquella aproximación que sacrifica precisión en la medida para aumentar la utilidad de los resultados de la evaluación para los directamente implicados en

el programa (1980a). Para Stake, la evaluación debe centrarse en las actividades del programa - no tanto en las intenciones -, responder a las necesidades de información de las partes implicadas y hacer referencia a las distintas perspectivas sobre los valores presentes al informar sobre el éxito o el fallo del programa (1980a).

Este modelo presenta las siguientes ventajas sobre la evaluación cuantitativa clásica:

1. permite que emerjan las variables importantes para el programa. Tradicionalmente se requiere al evaluador que especifique qué tratamientos y variables son más importantes. Stake entiende que ello no puede hacerse sin haber observado el programa o sin un cierto grado de flexibilidad para adaptar cambios en el programa.
2. facilita los cambios en el programa. Para Stake, las evaluaciones desarrolladas en relación con la reforma educativa en los Estados Unidos, han tenido un efecto desilusionador cuando la evaluación debería impulsar los cambios y las reformas. “El evaluador hace su mejor contribución cuando ayuda a descubrir ideas, respuestas, soluciones que muchas veces están en la mente de los que participan de una forma u otra en el programa (Stake,1975a, p.36).
3. devuelve el control al nivel local. Mientras que el modelo de evaluación clásico está dando el mensaje contrario “eres incompetente, necesitas ser dirigido”(1980b, p.161), Stake anima a los evaluadores a resistirse a aceptar los objetivos y criterios de las administraciones estatal y federal y a evitar que la evaluación sirva para dar mayor poder a determinados *stakeholders* (1979, p. 56).

Las recomendaciones a nivel metodológico pueden resumirse en las siguientes ideas:

1. dedicar mucho tiempo a la observación del programa, reunir valoraciones, ver qué necesidades tienen los clientes y preparar comunicaciones de carácter informal;
2. dejar que las cuestiones emerjan y cambien durante la evaluación. El contacto intensivo con el programa permite descubrir las cuestiones que van a estructurar las discusiones con los clientes;
3. cuidar la comunicación con las personas interesadas en la evaluación. Los resultados deberían expresarse de una forma natural que permita a su audiencia asimilar la información y comprender su significado (1980a, p. 83).

Stake muestra una clara preferencia por los estudios de caso. “El caso no tiene que ser una persona o un proyecto. Puede ser cualquier sistema con unos límites claros” (1978, p.7) Para Stake, un observador es el mejor instrumento para valorar muchas cuestiones que aparecen en una evaluación al poder ejercer su juicio crítico con mayor capacidad de apreciación de la realidad y de los resultados de un programa. La observación directa del caso es pues el elemento crucial de la evaluación. Para ello, los observadores pueden utilizar entrevistas, observación, examen de archivos y documentos, entre otros instrumentos metodológicos (1980a).

Para Stake, el objetivo de la evaluación es mejorar la práctica diaria del programa objeto de estudio y no producir nuevos conocimientos formales que puedan ser difundidos. En este contexto, el estudio de caso es relevante dado que puede generar un proceso de aprendizaje que se basa en el conocimiento tácito y en la asociación de ideas permitiendo “comprender las cosas de una forma distinta, ver nuevos significados y nuevas aplicaciones de lo antiguo” (Stake 1978, p.6). La base para la llamada “generalización natural” la pone el lector del caso al reconocer aspectos esencialmente similares a los de su experiencia (1978, p.7). El lector tiene la capacidad de reconocer lo que es relevante e irrelevante para sus circunstancias particulares. Stake reserva, pues, un papel importante a la experiencia y a las intuiciones de los profesionales interesados. La responsabilidad del investigador es proporcionar los datos contextuales que faciliten este poder intuitivo del lector. Por otra parte, el estudio de caso proporciona información que puede ser interpretada de distintas maneras sin favorecer ningún punto de vista por encima de los demás.

Los estudios de caso son especialmente adecuados para el modelo de evaluación de Stake. Sin embargo presentan algunas dificultades que el mismo Stake comenta:

El problema de la validez de un estudio de caso. La comprobación de la validez se puede conseguir con la triangulación de distintas investigaciones, intentando llegar a un mismo resultado por tres vías independientes. Esta idea es aplicable a un mismo estudio de caso o a distintos estudios de caso para un mismo programa en distintos puntos geográficos. Por otra parte, Stake entiende que el método de la “generalización natural” tiene la ventaja de que los lectores de un caso pueden participar en la determinación de su validez (Stake y Easley, 1978,:27).

El problema de la especificidad de los estudios de caso. Stake entiende, que los estudios de caso no son la alternativa mas adecuada cuando se requiere un cierto nivel de abstracción, o se pretende formular leyes generales y desarrollar teorías científicas. Aún así ,reivindica la utilidad de los estudios de caso para la función de exploración que toda actividad científica requiere (1978, p.7).

#### 6. Cronbach. Innovación metodológica y capacidad de generalización

Cronbach escribe sobre evaluación de programas después de haber dedicado una parte importante de su vida a temas metodológicos. Son importantes sus trabajos sobre fiabilidad (1951), validez interna (1955) y generalización (Cronbach y Snow, 1977). A mitad de los años 70, su visión de la teoría y métodos en ciencias sociales había evolucionado considerablemente. Adopta el termino *social inquiry*, en lugar de ciencias sociales para poner de relieve la necesidad de métodos interpretativos y descriptivos que puedan complementar los métodos tradicionales de muestreo, experimentación y análisis cuantitativo. Basándose en su propia experiencia, Cronbach ve la necesidad de descubrir nuevos métodos que se adapten a la naturaleza compleja y multivariable del contexto social.



En *Toward Reform of Program Evaluation* (1980), Cronbach y sus colaboradores de la Universidad de Stanford utilizan investigación empírica y teoría del análisis de políticas y de la evaluación de programas para presentar una teoría de la evaluación que pone el énfasis en la utilización de los resultados a corto plazo, tiene en cuenta el contexto político y administrativo caracterizado por la pluralidad de *satkeholders* y persigue satisfacer una necesidad de conocimiento que pueda ser extrapolado a otras situaciones.

Cronbach y otros (1980) entienden que buena parte de los problemas de uso de las evaluaciones provienen de modelo racional en que se fundamentan. Para estos autores, normalmente, a) no se toman decisiones de un forma clara en un momento determinado; b) hay más de un decisor; c) la política es más importante que la información sobre cómo optimizar las decisiones; d) no se toman decisiones del tipo si/no sobre la continuidad de un programa; e) no hay tiempo para prestar atención a todas las fases de una investigación; f) los informes llegan tarde; g) los decisores que piden un estudio y los que reciben sus resultados son distintos; h) los informes no están pensados para un gestor; i) los resultados de un estudio no se utilizan instrumentalmente para modificar un programa... En cambio, las políticas sociales son adoptadas cuando tienen un apoyo político suficiente, hasta el punto de poder tener una influencia relevante en votos. Los objetivos tienen que ser lo bastante vagos para poder acomodar múltiples intereses.

Todo ello sugiere la necesidad de incorporar el contexto político y administrativo a la teoría de la evaluación. Este entorno puede variar de forma considerable si se examinan las posibilidades que resultan de combinar actores y etapas de desarrollo de un programa. Concretamente, Cronbach y otros hablan de cinco grupos que influyen en el proceso de dar forma a una política (*policy shaping*): los responsables a nivel de formulación de la política (en el ejecutivo o en el legislativo), los responsables a nivel de programa (“los que tienen el nombre del programa en su puerta”), los prestadores de servicios a nivel local, los clientes y beneficiarios del programa, y los académicos, periodistas y otros grupos que interpretan o difunden información sobre el mismo. A su vez, distinguen entre tres etapas: una etapa inicial de probar nuevas ideas, una etapa de demostración en la que se maximizan las posibilidades de éxito, y una tercera etapa en la que el programa funciona en condiciones normales.

La idea de *policy shaping* se basa en la teoría del cambio social gradual. De acuerdo con esta concepción, Cronbach y otros rechazan que la evaluación sirva directamente para tomar decisiones o para controlar el rendimiento de un gestor. La evaluación sirve para “iluminar” las decisiones sobre distintas alternativas posibles, clarificando las cuestiones que interesan a las partes afectadas. La idea principal es entender el problema y el programa, el contexto político y, sobre todo, que procesos contribuyen a los resultados esperados. Esto último es particularmente importante para que estos procesos sean luego extrapolables a otras situaciones. Por ello, la teoría de Cronbach otorga un papel central a la validez externa.

Posteriormente, en *Designing Evaluations of Educational and Social Programs* (1982a), Cronbach ataca directamente los postulados de Campbell y sus colaboradores al defender

un nuevo concepto de validez que tiene que ver con la capacidad de producir conocimientos que trasciendan una situación específica. Para Cronbach,

1. la validez no es una propiedad del diseño de una investigación sino una propiedad de sus conclusiones, que tiene que ver con su credibilidad para las personas que podrían utilizar dichas conclusiones;
2. un tratamiento no se puede diferenciar de las unidades, observaciones y características contextuales en las que se presta y de las que su réplica depende;
3. existe una necesidad de extrapolar, esto es, de transferir los resultados de una investigación más allá del contexto en que se generan.

En cuanto al diseño de la evaluación, la propuesta de Cronbach se basa en tres principios:

1. no hacer un solo estudio de evaluación muy costoso sino varios estudios pequeños que tengan una única racionalidad;
2. no dedicar muchos recursos a responder con precisión unas pocas cuestiones específicas, sino dar respuestas menos claras a una amplia gama de cuestiones;
3. priorizar en función de la relevancia de las cuestiones para la comunidad de actores asociada a la política y del grado de incertidumbre en el conocimiento sobre los temas propuestos.

El papel del evaluador en este contexto es el de un educador independiente, no partisano, que informa a las partes, transmitiendo el conocimiento existente y enseñando a pensar críticamente (1980, p.67).

#### 7. Rossi. Especificación de modelos e integración teórica.

Rossi, un importante sociólogo que ha trabajado diversos aspectos de la metodología de las ciencias sociales, es una figura clave de la evaluación de programas tanto por sus trabajos de investigación aplicada (Berk et al. 1981, Rossi y Lyall, 1978), como por su trabajo conceptual de integración de las teorías desarrolladas por otros autores. Rossi es conocido por el libro de texto que escribió junto con Freeman y que ha sido uno de las principales referencias para la formación en evaluación de programas. Rossi presenta su aproximación integradora en base a tres conceptos: la idea de evaluación total, - *comprehensive evaluation* -, la evaluación a la medida, - *taylored evaluation* - y la evaluación sobre modelo teórico - *theory-driven evaluation* -.

Frente a la dispersión sobre las cuestiones a las que una evaluación debería responder, Rossi formula la idea de una evaluación que agrupa tres tipos de actividades: el análisis de la conceptualización y del diseño de una intervención - es preciso documentar la necesidad de un programa para planificar, llevar a término y evaluar acciones sociales -; el control de la implementación de un programa - el desarrollo de un simple sistema de indicadores para el control de actividades puede ser tan a más importante que producir información sobre su impacto -; y la evaluación de la utilidad del programa - salvo que los programas tengan un impacto demostrable, será difícil defender su continuidad -.

Dada la imposibilidad práctica de llevar a cabo una evaluación que desarrolle simultáneamente estas tres actividades, Rossi propone la idea de evaluación a la medida, con el objeto de adaptar cada evaluación al programa objeto de estudio. Rossi distingue entre programas nuevos, reformas o modificaciones de programas que ya funcionan y programas estables en pleno funcionamiento. Para cada estadio o nivel de desarrollo será preciso determinar las cuestiones y los métodos más adecuados. Así por ejemplo, en el caso de un programa nuevo, deberían dedicarse pocos recursos a estimar efectos. Únicamente tendría sentido dedicar esfuerzos a analizar los resultados de programas similares o a recoger las opiniones de expertos sobre los efectos que cabe esperar.

Rossi utiliza otro concepto, *theory-driven evaluation*, para reclamar mayor atención al desarrollo de modelos teóricos de las intervenciones sociales. En su trabajo con Chen, estos autores señalan la falta de atención en la literatura a la importante tarea de comprensión de las políticas sociales: “lo que es realmente atractivo de los experimentos controlados es que no es necesario entender como funciona un programa social para entender sus efectos netos...” (1983, p.284). Pero si los resultados del experimento no muestran efectos, no se sabe si el fallo es debido a que el programa estaba construido sobre fundamentos conceptuales pobres o a que el tratamiento era claramente insuficiente para producir efectos, o a que se han producido fallos en la implementación.

Rossi defiende la necesidad de especificar modelos sobre como cabe esperar que funcione un programa antes de evaluarlo. Si bien Rossi es claramente experimentalista en su primera época, posteriormente combina esta aproximación con la modelización estructural con el objeto de obtener información sobre los inputs, procesos intermedios y productos de los programas. Este esfuerzo por modelizar teóricamente la intervención social le permite integrar las aproximaciones dispares de Campbell y Cronbach a la cuestión de la validez. Para Rossi, se pueden tratar los problemas de validez interna especificando un modelo que tenga en cuenta las relaciones entre variables de la forma mas completa posible. La especificación de un modelo facilita asimismo la generalización en la medida que fuerza al investigador a definir el sistema en el que ha de operar el programa, las características contextuales y las posibles interacciones (Chen y Rossi, 1987).

En *Evaluation. A Systematic Approach*, Rossi y Freeman (1985) incorporan algunas consideraciones sobre la práctica de la evaluación.

Respecto a la cuestión sobre los stakeholders de un programa y la perspectiva que debería adoptar el evaluador, Rossi y Freeman otorgan una especial atención a los que están más implicados en la toma de decisiones. Sin embargo, el evaluador tiene que ser muy claro en lo que concierne a desde qué perspectiva se hace una evaluación y debe reconocer de forma explícita la existencia de otras perspectivas. Por otra parte, no debe incorporar los sesgos del que financia la evaluación, siendo esta una cuestión moral o de ética profesional.

Rossi prefiere las técnicas cuantitativas. Los métodos cualitativos suelen ser caros, difíciles de utilizar e imprecisos, según Rossi. Sin embargo, recomienda utilizar técnicas cualitativas para el análisis de la conceptualización, el diseño y para el control de las operaciones de un

programa. En este caso, es partidario de estructurar de alguna forma la recogida de datos para poder presentar información de forma resumida y que permita su análisis.

Toda la discusión de la evaluación en *Evaluation. A Systematic Approach* se organiza en base a la distinción entre programas innovadores, programas estables en funcionamiento y programas objeto de reformas y cambios. Al igual que Campbell, Rossi otorga mayor atención a los programas nuevos. La situación ideal para Rossi es la de trabajar con los responsables del programa optimizando las probabilidades de obtener una buena intervención y una buena evaluación. Respecto a este tipo de programas y evaluaciones, Rossi dedica una especial atención a las siguientes cuestiones:

1. La importancia de trabajar con objetivos específicos. Rossi discute distintas propuestas de otros autores -como los análisis de evaluabilidad de Wholey - en las que la idea principal es, sobre la base del acuerdo de los *stakeholders*, decidir criterios y niveles aproximados de éxito/fracaso. También recoge la posibilidad de incluir objetivos aceptados en la literatura del campo substantivo correspondiente.
2. Los modelos de impacto. Rossi distingue entre una hipótesis causal, una hipótesis de intervención y una hipótesis de acción. La primera se refiere a la relación causal, la segunda a como la intervención pretende afectar esta relación, y la tercera, a la posibilidad que la intervención no esté necesariamente relacionada con el resultado deseado, incluso en el caso que éste se produzca.
3. El análisis de la forma de prestación del servicio. Para Rossi, es crucial examinar si el programa se ha llevado a la práctica correctamente. Esto supone, analizar el problema y la población objetivo, los servicios prestados y sus componentes específicos, las características del personal, el proceso de selección de los beneficiarios y las formas de acceso al programa.
4. La posibilidad de experimentar. Lo que es más factible cuando se trata de programas nuevos, no extensivos a toda la población potencial sino únicamente a determinados grupos de beneficiarios, lo que permite construir grupos de control y establecer comparaciones. Al igual que Campbell, Rossi prefiere los experimentos para la evaluación de impacto. Sin embargo, recomienda seleccionar el mejor diseño posible, en función de las condiciones y circunstancias específicas de cada caso. “El evaluador debería escoger el mejor diseño posible, teniendo en cuenta que sea practicable y factible..., los recursos disponibles y los conocimientos del evaluador” (1985, p. 167).

Respecto al problema de la utilización de las evaluaciones, Rossi defiende una utilización instrumental de sus resultados. Para que ello sea posible, cree necesario que los evaluadores entiendan la forma de trabajar de los decisores, lo que implica, primero, presentar los resultados a tiempo, cuando son necesarios y de forma breve; segundo, ser sensible a las posiciones de todas las partes; y, tercero, planificar su difusión y utilización. También ve la posibilidad de una utilización conceptual o de una utilización política, para atacar o defender determinadas posiciones (1985, p.188).

### III. El marco para la práctica de la evaluación

La evolución teórica de la evaluación es fruto de la práctica profesional durante los últimos veinticinco años. La teoría, a su vez, es útil como marco de referencia para una práctica profesional que a menudo precisa de orientación en su aproximación a los problemas concretos con los que se enfrenta. De la observación del cuadro 1 se podría concluir que hay una considerable diversidad de aproximaciones en la disciplina. Sin embargo, también se produce una cierta convergencia al irse aceptando algunos principios generales que pueden servir de guía para la práctica de la evaluación.

Así, un profesional que se enfrenta al reto de una evaluación, tiene que optar entre múltiples alternativas, cuya argumentación teórica parece irrefutable. Pero, a la vez, tiene algunos criterios que puede seguir con una cierta seguridad cuando se plantea cuestiones críticas del tipo: ¿debería hacer esta evaluación?, ¿en qué debería centrarse?, ¿cómo se puede hacer?, ¿desde qué perspectiva?, ¿cómo se puede facilitar su utilización posterior?

Si la cuestión es *qué* evaluar, la teoría ofrece distintas posibilidades que van desde los efectos hasta las necesidades sociales, pasando por los procesos intermedios y los análisis coste-eficacia. La teoría ofrece asimismo distintos actores que pueden responder esta cuestión: directivos públicos, clientes, profesionales... las instituciones que financian la evaluación. Existe pues una variedad de respuestas posibles a la pregunta *quien decide o para quien* se hace la evaluación.

El contrapunto a esta amplia gama de posibilidades está en algunos principios que indican al evaluador como se relacionan estas variables. Así, parece claro que, cada actor tiene preocupaciones distintas en función del lugar, la responsabilidad o el interés que tiene en el programa. Sin embargo, independientemente de quien sea el cliente del estudio, el evaluador ha de tener en cuenta los valores u objetivos del conjunto de *stakeholders*, incluyendo a afectados directamente o indirectamente por el programa, y a decisores con poder efectivo. Por otra parte, el evaluador deberá responder a las necesidades de información de los comanditarios de la evaluación si quiere facilitar su posterior utilización. A su vez, la opción sobre el objeto de la evaluación estará determinada por la etapa de desarrollo en la que se encuentra el programa, la naturaleza del programa, la información existente sobre el mismo, y los recursos - humanos y materiales - invertidos en la evaluación.

Si la pregunta que se plantea el evaluador es *cómo* hacer su estudio, las respuestas ilustran una división clara a nivel metodológico. La gama de posibilidades incluye: métodos de tipo experimental, métodos cuantitativos clásicos (estadística descriptiva, análisis inferencial, correlación, regresión, sondeos), métodos de control de gestión (sobre la base de sistemas de indicadores), métodos cualitativos clásicos (observación, entrevistas, análisis de documentos) y estudios de caso.

Sin embargo, como se desprende de la teoría, la finalidad y el objeto de la evaluación determinan considerablemente las opciones metodológicas. Así, por ejemplo, existe un cierto acuerdo en que los estudios de caso no son la mejor opción para el análisis de efectos

y de inferencias causales o en que cuestiones de conceptualización y diseño de una intervención en un ámbito social sobre el que el conocimiento está poco estructurado requieren una aproximación más cualitativa.

Cuadro 1. Modelos teóricos en evaluación de programas

	<b>SCRIVEN</b>	<b>CAMPBELL</b>	<b>WEISS</b>	<b>WHOLEY</b>	<b>STAKE</b>	<b>CRONBACH</b>	<b>ROSSI</b>
<b>Nuevos términos</b>	<b>Ciencia de valoración</b>	<b>Sociedad Experimental</b>	<i>Enlightening Model</i>	<b>Análisis de Evaluabilidad</b>	<i>Responsive, Naturalistic Model</i>	<b>Analisis de UTOS (unidades, tratamientos, observación, contexto)</b>	<i>Theory-Driven Model</i>
<b>Autores con una visión similar</b>		<i>Boruch, Cook, Riecken, Rossi</i>	<i>Cronbach, Chelimsky, Palumbo, Wholey</i>	<i>Abramson, Bellavita, Horst, Nay, Scanlon</i>	<i>Guba, Lincoln, Patton</i>		<i>Berk, Chen, Freeman, Wright</i>
<b>Concepto</b>	Juicio de valor	Producir información para la toma de decisiones	Clarificar cuestiones, “iluminar” a decisores y afectados	Producir información para la gestión	Producir información práctica y útil para los miembros del programa	Clarificar cuestiones, sobre: -problema, -programa -contexto	Producir información para la toma de decisiones
<b>Interés prioritario</b>	Interés general	Responsables del programa	Legislador y directivos	Legislador y directivos	Beneficiarios del programa, profesionales	Legislador y directivos	Responsables del programa
<b>Rol del evaluador</b>	Juez	Metodólogo	Educador	Agente de cambio	Prestador de un servicio	Educador, Difusor	Metodólogo,
<b>Objeto</b>	Efectos	Efectos Rel. causales	Efectos, procesos, contexto	Efectos, actividades, productos	Actividades	Efectos, procesos, contexto	concepto, implementación, eficacia
<b>Fuente de criterios para determinar eficacia</b>	Necesidad Social	objetivos,	objetivos, <i>stakeholder</i> , documentación	<i>stakeholder</i>	clientes, técnicos	<i>stakeholder</i>	<i>stakeholder</i> , literatura científica

<b>Métodos</b>	Lógica experimental	Experimento Cuasi experimento	Métodos cuantitativos Métodos cualitativos	Análisis de evaluabilidad Sistemas de indicadores Evaluación rápida	Métodos cualitativos Estudios de caso	Métodos cuantitativos Métodos cualitativos	Experimento Cuasi experimento Métodos cualitativos, excepto relaciones causales
<b>Perspectiva sobre utilización</b>	Instrumental	Instrumental	Conceptual Formulación de políticas	Instrumental	Instrumental	Conceptual Comprensión de la realidad social	Instrumental Conceptual y Política. Formulación de teoría
<b>Oportunidad de la evaluación</b>	Criterio coste-eficacia	Voluntad de reforma, superada la etapa de iniciación	En función de las tres I: - intereses, - información disponible, - ideología	Criterio coste-eficacia	Util para los miembros del programa	En función de la relevancia del estudio, y de la falta de información	Programas innovadores fase de prueba piloto
<b>Organización de la evaluación</b>	Operación externa	operación externa	Implicar al personal, no impedir la prestación del servicio	Evaluación Parte del sistema de gestión por resultados	Contacto permanente con el programa	Obtener la cooperación del programa	Adaptarse al tiempo y forma de trabajar de los gestores



Otra cuestión que puede tener respuestas plurales es la que se plantea la contribución que puede realmente esperarse de una evaluación. Aquí, los dos extremos estarían representados por las siguientes posiciones. Por una parte, se puede entender que la evaluación únicamente tiene sentido en cuanto resulta útil de forma inmediata, a corto plazo, de ahí, su financiación. Por otra parte, la evaluación será realmente útil en cuanto contribuya a producir un cambio en la forma de concebir una determinada realidad social y las posibilidades de intervención sobre ésta desde las instituciones públicas.

Nuevamente, existe una posibilidad de integración de estas dos posturas sobre la base de diferenciar dos realidades diversas: la evaluación en tanto que servicio a los profesionales o directivos de un programa concreto, y la evaluación en tanto que oportunidad para la reflexión sobre una determinada forma de tratar una realidad social problemática. En la primera, no hay una preocupación teórica fundamental que si está presente en la segunda, donde la información específica sobre el detalle del funcionamiento del programa puede ser menos relevante. Normalmente, cabría esperar que los evaluadores académicos tengan un mayor interés por intervenir en las evaluaciones del segundo tipo, mientras que los consultores profesionales lo harán en las del primer tipo.

Otra pregunta que puede hacerse el evaluador es *cuando* tiene sentido hacer una evaluación. Las respuestas posibles van desde las prescripciones técnicas en función de los métodos empleados a consideraciones sobre el contexto político (intereses, ideología, propósitos ocultos) y administrativo (urgencia de la decisión, percepción de la evaluación por el personal afectado, medida en que la evaluación puede afectar la prestación regular del servicio) o sobre los recursos que requiere.

De hecho, todas estas consideraciones son perfectamente compatibles. Una característica que se desprende del estudio de la teoría de la evaluación es que se pasa de una consideración del programa en un vacío social, a la progresiva incorporación del contexto y de la preocupación por adaptar la evaluación a un sistema político plural y descentralizado. La práctica de la evaluación no puede por tanto obviar que la evaluación es política y que sus resultados son una información más en el proceso político que tiene que competir con informaciones que tienen su acceso por otras vías.

Aún otra cuestión que a menudo se plantean directivos y evaluadores es como organizar la evaluación. Las prescripciones van desde la organización de la evaluación como una operación externa a la integración de la evaluación en el equipo que desarrolla el programa, pasando por la lógica formación de equipos mixtos. Cada una de estas opciones tiene sus ventajas y sus inconvenientes, mientras que no siempre se disponen de recursos suficientes para montar el mejor dispositivo de evaluación.

Sin embargo, la teoría es clara respecto a esta cuestión. Raramente se puede hacer una evaluación desde un despacho, sin el contacto con la realidad del programa. Es necesario pues mantener el contacto, alimentar el interés en la evaluación e implicar a las personas claves en las principales decisiones que se toman durante el proceso evaluador para facilitar su posterior utilización.

#### **IV. Conclusiones**

El repaso de las teorías de siete autores claves en evaluación de programas, siguiendo el trabajo de Shadish, Cook y Leviton (1991), ilustra la evolución de la disciplina y su progresiva estructuración sobre la base de otras ciencias sociales y de los principios extraídos de la propia experiencia en evaluación. Crane (1988) utiliza la metáfora de una “zona de pastos común,” para poner de relieve el carácter interdisciplinario de la evaluación de programas. Cronbach y sus colaboradores (1980) hablan de una “industria de servicios” donde importa menos la afiliación que la capacidad para resolver problemas, poniendo así de relieve su carácter práctico.

Los teóricos de la evaluación han desarrollado de forma intensiva los aspectos metodológicos, dada la necesidad de dar respuestas creíbles a los problemas planteados por la administración. También han dedicado una parte importante de su trabajo a desarrollar aspectos de proceso, evolucionando desde una concepción de la evaluación como ciencia a una nueva concepción de la evaluación como arte, en el sentido de habilidad para construir una investigación de calidad que resulte útil en un determinado contexto político-administrativo. Todo ello sin olvidar que evaluar implica necesariamente la idea de juicio de valor y por tanto la adopción de unos determinados principios que luego son traducidos a criterios de valoración.

La evolución teórica de la evaluación de programas es fruto de la práctica profesional. Por ello, se ha producido una diversificación considerable. Se ha pasado del estudio exclusivo de los efectos, a la preocupación por examinar la aplicación concreta de los programas y los procesos causales que median en la producción de los impactos deseados. La confianza exclusiva en aproximaciones de tipo cuantitativo-experimental ha dado paso a la inclusión de los métodos cualitativos y de los estudios de caso. Los legisladores y directivos públicos fueron en una primera etapa la única fuente de las cuestiones y de los criterios de evaluación, mientras que progresivamente, otras partes interesadas han sido incluidas en esta consideración. La utilización efectiva de las evaluaciones, un tema inicialmente ignorado, ha pasado a ser una de las principales preocupaciones en la disciplina.

La diversidad en las aproximaciones no es contradictoria con un proceso de integración de la teoría, que aun queda por hacer, pero que es evidente cuando se analiza la literatura. La evaluación no es pues una disciplina a teórica ni la teoría de la evaluación es un conjunto de proposiciones de escasa utilidad para la práctica de la disciplina. Al contrario, la evaluación tiene en la teoría emergente un conjunto de proposiciones que pueden dar respuestas útiles a las cuestiones críticas que plantea la práctica de la evaluación y de las que depende que ésta sea realmente útil para la reforma de las intervenciones públicas y la formulación de nuevas políticas.

## V. Bibliografía

Abramson, M.A., y Wholey, J.S. (1981). Organization and Management of the Evaluation Function in a Multilevel Organization. In R.J. Wooldridge (Ed.), *Evaluation of Complex Systems* (pp.31-48). San Francisco:Jossey-Bass.

Alvira Martín, F. (1991) *Metodología de la Evaluación de Programas*. Madrid, CIS.

Ballart, X. (1992) *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*. Madrid. Ministerio Administraciones Públicas.

Ballart, X. (1994) “Clientes de un servicio cultural: una experiencia de evaluación con ‘focus groups’”, *Gestión y Análisis de Políticas Públicas*, Num. 1, 1994.

Ballart, X. y Riba, C. (1996) “Impact of legislation requiring moped and motorbike riders to wear helmets”, *Evaluation and Program Planning*, núm.1, 1996.

Berk, R.A., Lenihan, K.J., & Rossi, P.H. (1981). Crime and Poverty: Some Experimental Evidence from Ex-Offenders. In H.E. Freeman & M.A. Solomon (Eds.), *Evaluation Studies Review Annual* (Vol. 6. pp 339-359). Beverly Hills, CA:Sage.

Brizius, J.A., Campbell, M.D. (1991), *Getting results: A Guide for Government Accountability*, Washington, DC. Council of Governors’ Policy Advisors.

Campbell, D.T. (1969). Reforms as Experiments. *American Psychologist*, 24, 409-429.

Campbell, D.T. (1970). Considering the Case Against Experimental Evaluations of Social Innovations. *Administrative Science Quarterly*, 15,110-113.

Campbell, D.T. (1982). Experiments as Arguments. In E.R. House, S. Mathison, J.A. Pearsol, & H. Preskill (Eds.), *Evaluation Studies Review Annual* (Vol. 7,pp. 117-127). Beverly Hills, CA:Sage.

Campbell, D.T. (1988). *Methodology and Epistemology for Social Science: Selected papers* (E.S. Overman, Ed.) Chicago: University of Chicago Press.

Campbell, D.T., & Boruch, R.F. (1975). Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects. In C.A. Bennett & A.A. Lumsdaine (Eds.), *Evaluation and Experiments:Some Critical issues in Assessing Social Programs* (pp. 195-296). New York: Academic Press.

Campbell, D.T., & Stanley, J.C. (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago:Rand McNally.

Chelmsky, E., (1987) "The Politics of Program Evaluation" en D.S. Cordray, H.S. Bloom y R.J. Light, *Evaluation Practice in Review*. San Francisco, Jossey-Bass.

Chen, H. & Rossi, P.H. (1983). Evaluating with Sense: The Theory-Driven Approach. *Evaluation Review*, 7, 283-302.

Chen, H. & Rossi, P.H. (1987). The Theory-Driven Approach to Validity. *Evaluation and Program Planning*, 10, 95-103.

Cook, T.D., D.T. Campbell (1979). *Quasiexperimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Crane, J.A. (1988). Evaluation as Scientific Research. *Evaluation Review*, 12, p. 467

Cronbach, L.J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.

Cronbach, L.J. (1982a). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.

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, CA: Jossey-Bass.

Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.

Cronbach, L.J. & Snow, R.E. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington.

Dilulio Jr, J.J. (1994) *Measuring Performance When There is No Bottom Line*, Center for Public Management Working Paper 94-2, Washington DC: The Brookings Institution.

Duran, P., Monnier, E., Smith, A. (1995)  
"Evaluation a la française", *Evaluation*, núm. 1 1995.

Guba, E.G. & Lincoln, Y.S. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*. San Francisco: Jossey-Bass.

Guba, E. G. & Lincoln, Y.S. (1989) *Fourth Generation Evaluation*, Beverly Hills, CA: Sage.

Hendricks, M. (1981). Service Delivery Assessment: Qualitative evaluations at the cabinet level. In N.L. Smith (Ed.) *Federal Efforts to Develop New Evaluation Methods* (pp.5-24). San Francisco: Jossey-Bass.

- Horst, P., Nay, J.N., Scanlon, J.W., & Wholey, J.S. (1974). Program Management and the Federal Evaluator. *Public Administration Review*, 34,300-308.
- House, E.R. (1980). *Evaluating with Validity*. Beverly Hills, CA:Sage.
- Lincoln, Y.S., Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA:Sage.
- Palumbo, D.J. (1987) *The Politics of Program Evaluation*, Beverly hills, CA: Sage.
- Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills, CA:Sage.
- Pollit, C. (1992). *Occasional Excursions: A Brief History of Policy Evaluation in the UK*. MEANS program, Bruselas: DGXVI CEE.
- Rossi, P.H., & Freeman, H.E. (1982, 1985, 1989). *Evaluation: A Systematic Approach*. Beverly Hills, CA:Sage.
- Rossi, P.H. & Lyall, K.C. (1978). An Overview Evaluation of the NIT Experiment. In T.D. Cook, M.L. DelRosario, K.M. Hennigan, M.M. Mark, & W.M.K. Trochim (Eds.), *Evaluation Studies Review Annual* (Vol. 3, pp.412-428). Beverly Hills, CA:Sage.
- Scriven, M. (1967). The Methodology of Evaluation. In R.W.Tyler, R.M.Gagne, & M.Scriven (Eds.), *Perspectives of Curriculum Evaluation* (pp.39-83). Chicago: Rand McNally.
- Scriven, M.(1972). The Methodology of Evaluation. In C.H. Weiss (Ed.), *Evaluating Action Programs: Readings in Social Action and Education* (pp.123-136). Boston: Allyn & Bacon.
- Scriven, M (1974). Evaluation Perspectives and Procedures. In J.W.Popham (Ed.), *Evaluation in Education: Current Application* (pp.3-93). Berkeley, CA:McCutchan.
- Scriven, M.(1980). *The Logic of Evaluation*. Inverness, CA: Edgepress.
- Scriven, M (1981). The Good News and the Bad News about Product Evaluation. *Evaluation News*,2,278-282.
- Scriven, M.(1983).Evaluation Ideologies. In G.F.Madaus, M. Scriven, & D.L.Stufflebeam (Eds.), *Evaluation Models:Viewpoints on Educational and Human Services Evaluation* (pp.229-260). Boston: Kluwer-Nijhoff.
- Shadish, W.R. Epstein, R. (1987). Patterns of Program Evaluation Practice. Among Members of the Evaluation Research Society and Evaluation Network. *Evaluation Review*, 11, p.555

- Shadish, W.R. Cook, T.D. Leviton. L.C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, Ca: Sage.
- Smith, N.L. (1981). Classic 1960s Articles in Educational Evaluation. *Evaluation and Program Planning*, 4, 177-183.
- Stake, R.E. (1975a). An Interview with Robert Stake on Responsive Evaluation. In R.E. Stake (Ed.), *Evaluating the Arts in Education: A Responsive Approach* (p. 33-38). Columbus, OH:Merrill.
- Stake, R.E. (1975b) To Evaluate an Arts Program. In R.E. Stake (Ed.). *Evaluating the Arts in Education: A Responsive Approach* (pp.13-31). Columbus, OH:Merrill.
- Stake, R.E. (1978). The Case Study Method in Social Inquiry. *Educational Researcher*, 7,5-8.
- Stake, R.E. (1979). Validating Representations: The Evaluator's Responsibility. In R.Perloff (Ed.), *Evaluator Intervention: Pros and Cons* (pp. 55-70). Beverly Hills, CA: Sage.
- Stake, R.E. (1980a). Program Evaluation, Particularly Responsive Evaluation. In W.B. Dockrell & D. Hamilton (Eds.), *Rethinking Educational Research* (pp. 72-87). London:Hodder & Stoughton.
- Stake, R.E. (1980b). Quality of Education and the Diminution of Local Control in Schools in the United States. In U.S. House of Representatives,Committee on Education and Labor, *Needs of Elementary and Secondary Education in the 1980's* (pp. 161-168). Washington, DC:U.S. Government Printing Office.
- Stake, R.E., & Gjerde,C. (1974). *An Evaluation of TCITY: The Twin City Institute for Talented Youth* (AERA Monograph Series in Curriculum Evaluation, No.7). Chicago: Rand McNally.
- Stake, R.E. & Easley, J.A. (1978). *Case Studies in Science Education*. Champaign. University of Illinois,Center for Instructional Research and Curriculum Evaluation.
- Weiss, C.H. (1972a). Evaluating Educational and Social Action programs. A Treeful of Owls. In C.H. Weiss (Ed.), *Evaluating Action Programs. Readings in Social Action and Education* (pp.3-27). Boston:Allyn & Bacon.
- Weiss, C.H. (1972b). *Evaluation Research:Methods for Assessing Program Effectiveness*. Englewood Cliffs, N. J:Prentice-Hall.
- Weiss, C.H. (1973a). The Politics of Impact Measurement. *Policy Studies Journal*, 1, 179-183.

Weiss, C.H. (1973b). Where Politics and Evaluation Research meet. *Evaluation*, 1, 37-45.

Weiss, C.H. (1980). Efforts at Bureaucratic Reform: What Have We Learned? In C.H. Weiss & A.H. Barton (Eds.) *Making Bureaucracies Work* (pp. 7-26). Beverly Hills, CA:Sage.

Weiss, C.H. (1983a). Toward the Future of Stakeholder Approaches in Evaluation. In A.S. Bryk (ed.), *Stakeholder-Based Evaluation* (pp. 83-96). San Francisco: Jossey-Bass.

Weiss, C.H. (1983b). Ideology, Interest, and Information: The Basis of Policy Decisions. In D. Callahan & B. Jennings (Eds.), *Ethics, the Social Sciences, and Policy Analysis* (pp. 213-245). New York: Plenum.

Weiss, C.H. (1987). Evaluating Social Programs: What Have We Learned? *Society*, 25(1), 40-45.

Weiss, C.H. & Bucuvalas, M.J. (1980). *Social Science Research and Decision-Making*. New York: Columbia University Press.

Weiss, R.S., & Rein, M. (1970). The Evaluation of Broad-Aim Programs: Experimental Design, its Difficulties, and an Alternative. *Administrative Science Quarterly*, 15, 97-109.

Wholey, J.S. (1979). *Evaluation: Promise and Performance*. Washington, DC: Urban Institute.

Wholey, J.S. (1983). *Evaluation and Effective Public Management*. Boston: Little Brown.

Wholey, J.S., Abramson, M.A. y Bellavita, C. (eds.) (1986) *Performance and Credibility. Developing Excellence in Public and Nonprofit Organizations*. Lexington, MA: Lexington.