

## EVALUACIÓN DE LOS RESULTADOS Y DEL IMPACTO

Pedro M. Apodaca  
Unidad de Evaluación  
Universidad del País Vasco

### I. ACLARANDO TÉRMINOS

Cuando hablamos de evaluación del impacto en realidad estamos refiriéndonos a la evaluación de los *efectos* producidos por un programa o intervención. Es decir, aquellos cambios habidos debidos a la citada intervención. Sin embargo, no todos los cambios ocurridos durante o después de un programa o intervención son debidos a la propia intervención. En efecto, numerosos fenómenos y factores acompañan la intervención provocando, de forma directa o indirecta, sus propios efectos en los destinatarios del programa.

Por lo tanto, debemos distinguir entre el conjunto de cambios/resultados habidos tras un programa o intervención que denominaremos *efectos brutos* y aquella parte que son efecto real de la propia intervención o programa y no de otros factores o circunstancias; *efectos netos*.

Cuando hablamos de los *resultados* de un programa nos referimos frecuentemente a los *efectos brutos* del mismo. Cuando hablamos de *evaluación del impacto* deberemos referirnos siempre a los *efectos netos*. Sin embargo, no es ésta la única clarificación diferencial que podemos encontrar al respecto.

Frecuentemente se utiliza el término *evaluación de resultados* como sinónimo de *evaluación del impacto*. Resulta difícil identificar matices diferenciales de ambos términos debido a que se utilizan profusamente como si fueran equivalentes. Una de las pocas aclaraciones terminológicas nos la proporciona Alvira (1991) para quien la *evaluación de los resultados* haría referencia al

intento de analizar los efectos que un programa o intervención tiene en los sujetos directamente destinatarios del mismo. En la evaluación del impacto se trataría de analizar los efectos del programa en un marco más amplio, en el conjunto de la población que, sin ser directamente destinataria del programa o servicio, es influida de alguna manera.

El libro de Schalock (1995) también introduce una particular estructura terminológica al distinguir tres tipos de *evaluación basada en los resultados*. Se trata del análisis de la *efectividad*, del análisis del *impacto* y del análisis *coste-beneficio*. De esta manera, la evaluación del impacto sería uno de los tipos de análisis en la evaluación basada en los resultados. En esta misma obra aclara que entiende por *impacto* la diferencia entre los *resultados* obtenidos por dos programas o intervenciones. Esto viene a ser equivalente a la formulación de Rossi y Freeman (1993) quienes entienden por *impacto* los efectos netos de un programa. Del resultado o efecto bruto descubierto habría que desglosar los efectos de los factores extraños y los efectos del diseño para extraer el efecto neto del programa.

En definitiva, podemos concluir que el término *evaluación del impacto* resulta más operativo e induce a menos confusiones que el término *evaluación de resultados*.

En un intento de acotar y clarificar podríamos definir el término «evaluación del impacto» como la denominación genérica para toda estrategia de evaluación que trate de analizar y verificar los *efectos* producidos por una intervención o programa concreto.

## 2. CAUSALIDAD Y EVALUACIÓN DEL IMPACTO

El problema de la causalidad sigue siendo uno de los grandes quebradores de cabeza en las Ciencias Sociales y Humanas. Existen diversas concepciones alternativas de la causalidad y diversos procedimientos o requisitos para su afirmación (Dendaluce, 1994). Esta situación de cierta confusión y de teorías alternativas es también vigente en la evaluación.

La evaluación del impacto, en su formulación tradicional (Rossi y Freeman, 1993), presupone que determinados resultados o productos aparecidos durante o después de un programa son efecto de la intervención. Desde un orientación experimental clásica, esta inferencia causal se garantizaría mediante la aleatorización y la manipulación. Sin embargo, la inferencia causal necesita también de generalización y no sólo de validez interna ya que la inducción de un caso particular a una regla universal no está justificada lógicamente.

En cualquier caso, la evaluación del impacto no puede proporcionar seguridad plena de que determinados resultados son efecto del programa. Únicamente puede proporcionar diversos grados de plausibilidad. Lo riguroso del diseño en que se fundamente la evaluación del impacto añade plausibilidad a la inferencia causal. Sin

embargo, no todos los tipos de diseños son viables puesto que en su elección deben tenerse en cuenta también consideraciones de tipo práctico, económico, ético, político, etc.

Así pues, desde los paradigmas clásicos de la investigación y de la evaluación de programas, podemos afirmar que la evaluación del impacto presupone causalidad. Esta causalidad se apoya en la aleatorización a todos los niveles y en la manipulación de la variable independiente o intervención. Sin embargo, desde otros paradigmas más participativos, la explicación formalizada de los fenómenos pierde relevancia frente a la comprensión de los mismos; el pensamiento empírico, técnico y racional frente al pensamiento simbólico, mitológico y mágico; la denotación frente a la connotación (Ander-Egg 1990). Bajos estos paradigmas, la idea clásica de causalidad y el interés por la demostración de los efectos del programa (impacto) no serían prioritarios en la evaluación.

Una de las formulaciones más clarificadoras en este sentido la podemos encontrar en la obra de Guba y Lincoln (1981, 1989) quienes aportan una buena síntesis de los principios ontológicos, epistemológicos y metodológicos que contraponen los paradigmas positivista-cuantitativo y naturalista-cualitativo. Bajo su paradigma *respondente-constructivista* postulan, a nivel *Ontológico*, que no existe una realidad objetiva y la ciencia misma es una construcción social. A nivel *Epistemológico* afirman que no cabe plantear una dualidad entre sujeto y objeto de la evaluación. Los resultados de la evaluación son el producto de las interacciones entre observadores y observados. A nivel *Metodológico* rechazan el enfoque clásico de control y manipulación (experimental) y lo sustituyen por un proceso dialéctico-hermenéutico que ha de dar cuenta de la interacción entre observador y observado. Un proceso continuo de interacción, análisis, crítica y reanálisis.

La obra de Shadish, Cook y Leviton (1991) contiene también un buen análisis crítico de los diferentes paradigmas o estrategias de construcción del conocimiento presentes en la evaluación de programas.

### **3. LA EVALUACIÓN DEL IMPACTO EN DIFERENTES MODELOS DE EVALUACIÓN**

La evaluación del impacto tiene una mayor relevancia en determinados modelos de evaluación. Los modelos de consecución de Metas (Tyler et al 1989, Tyler 1991) establecen como criterio principal de mérito o valor la consecución de los objetivos explícitos y planificados del programa. Es decir, el análisis de la efectividad del programa es quizás el elemento principal de este modelo. Dentro de estos modelos existen numerosas variantes y, entre ellas, podemos destacar la propuesta de Vedung (1997) que denomina *evaluación de los efectos colaterales* en la que se señala la importancia de analizar también todos aquellos efectos no planificados o perseguidos por el programa.

Otros modelos de evaluación acentúan la importancia de los criterios intrínsecos en la formulación de los juicios de valor. Estos criterios se refieren a características del programa, de la institución, etc. Suelen ser denominados también modelos orientados

al juicio, evaluación colegiada, etc. Aún admitiendo grandes variantes de esta orientación, podemos decir que la caracteriza una escasa atención a los efectos, resultados o productos del programa.

Por el contrario, otros modelos de evaluación tienen en cuenta criterios de mérito de tipo extrínseco, es decir, del impacto real del programa. Dentro de estos modelos existen orientaciones diversas como las de Scriven (1983, 1991), Stake (1983, 1991), Cronbach (1982, 1983), Rossi y Freeman (1989, 1993), Stufflebeam y Shinkfield (1987).

Otros modelos han tomado orientaciones más radicales al establecer como criterio único de valor o mérito de un programa la valoración que del mismo hacen sus actores principales. Es lo que podríamos llamar el enfoque del participante o la evaluación como negociación que Guba y Lincoln (1981, 1989) denominan evaluación de cuarta generación o respondiente-constructivista. Bajo este paradigma la evaluación del impacto del programa queda diluida en una valoración interpretativa y constructivista de los implicados, los cuales pueden estar más interesados en los aspectos procesuales y no necesitar apoyar sus juicios en evidencias sobre los efectos reales de un programa.

TABLA 1  
DIFERENCIAS ENTRE EVALUACIÓN TRADICIONAL Y PARTICIPATIVA

	Evaluación tradicional	Evaluación participativa
<b>Quién</b>	Expertos externos	Comunidad, facilitador, gestores,...
<b>Qué</b>	Indicadores preestablecidos de éxito, de productividad	Los participantes identifican sus propios indicadores de éxito, los cuales pueden incluir la productividad
<b>Cómo</b>	Centrado en la objetividad científica,...	Autoevaluación y métodos simples adaptados a la cultura del grupo
<b>Cuando</b>	Frecuentemente al finalizar y algunas veces durante el proceso	Se mezclan la monitorización y la evaluación realizándose pequeñas y frecuentes evaluaciones.
<b>Porqué</b>	Rendición de cuentas, normalmente sumativa	Otorgar capacidad al grupo participante para iniciar, controlar o modificar su acción
<b>Generalización</b>	Objetivo básico de la evaluación. La aleatorización puede apoyar la validez externa de población. El control y manipulación pueden limitar la validez ecológica	El objetivo está centrado en el cambio y la mejora en el entorno de aplicación más que en su generalización a otros colectivos o situaciones.
<b>Validez Interna</b>	Se basa en la manipulación, la aleatorización y el control; siguiendo el enfoque experimental clásico.	Se basa en la triangulación de información y en el consenso como validez intersubjetiva.
<b>Eval. del Impacto</b>	Impacto como efectos netos de un programa desglosados los efectos de los factores extraños y del diseño.	El impacto se valora a partir de la percepción social o comunitaria de los efectos/utilidad del programa o intervención.

Una buena clasificación de los diferentes modelos de evaluación según diversos criterios podemos encontrarla en el trabajo de De Miguel (1998). Sin embargo, para los fines de este trabajo nos contentaremos con proponer una simplificación, probablemente excesiva, contraponiendo la evaluación tradicional a la evaluación participativa. Esta contraposición es muy clarificadora y podemos encontrarla tanto en manuales de iniciación (Alvira, 1991) como en obras de mayor alcance (Vedung, 1997, Rebien, 1997). Tomando algunos elementos de estas obras y añadiendo otros de elaboración propia, la Tabla 1 puede mostrar algunas de las características diferenciales de estos modelos:

El trabajo de De Miguel (1995) subraya la necesidad de tomar en consideración la participación activa de los implicados, y del *ciudadano* en general, en la evaluación de las políticas públicas, así como que ésta debe llevarse a cabo por equipos externos a la propia Administración. Señala asimismo que '*... el ejercicio del control democrático sobre las políticas públicas no se ejerce plenamente mientras que no exista una participación generalizada de la audiencia relativa a la ciudadanía en los comités de evaluación externa...*' (op. cit., p. 213).

Los enfoques que otorgan un gran protagonismo a los actores principales del programa ofrecen tres tipos de ventajas. Por un lado, aprovechan el conocimiento de la realidad que sólo pueden aportar los implicados. Por otro lado, garantizan el compromiso de éstos con la puesta en marcha de las acciones de mejora fruto del proceso de evaluación. Finalmente, dan cumplimiento al mandato de participación ciudadana propio de las democracias representativas. Las propuestas y postulados de Floc'Hlay y Plottu (1998) bajo la denominación de evaluación democrática van en esta línea. Fetterman, Kaftarian y Wandersman (1996) prefieren hablar de evaluación por delegación (empowerment).

Esta contraposición entre enfoque tradicional y participativo se modera rápidamente al abordar evaluaciones concretas. Así, si tomamos como referencia el *modelo respondiente-constructivista* de Guba y Lincoln entenderemos la evaluación como un proceso negociador, interpretativo y hermenéutico. Bajo el paradigma tradicional, el criterio fundamental de verdad es la correspondencia con los hechos, con los datos. Por el contrario, bajo este paradigma el criterio fundamental de verdad es el *consenso*. La verdad y lo real no es sino el consenso entre las audiencias. Enfatiza por tanto los juicios adecuadamente contextualizados y compartidos frente a los análisis más técnicos y supuestamente objetivos realizados por expertos. Ahora bien, esta orientación presenta dos importantes limitaciones: operativa y democrática. Por el lado operativo, en el contexto de grandes organizaciones o amplios programas de intervención, la búsqueda del consenso puede llegar a resultar inviable. La búsqueda del consenso en pequeños grupos que analizan una realidad limitada y un contexto próximo puede ser factible. Sin embargo, cuando el conjunto de aspectos a analizar es amplio y cuando las dimensiones del programa hacen difícil o imposible conocer directamente la realidad, se hacen precisos procedimientos y herramientas para la recogida y sistematización de la información. El uso de indicadores (tanto cualitativos como cuantitativos) se convierte en herramienta fundamental para acelerar el consenso y dinamizar y simplificar el trabajo de las comisiones. Esta información de base podrá ser adecuadamente

valorada e interpretada por las comisiones en relación al contexto del programa y a los intereses y preocupaciones de sus integrantes. Por otro lado, de esta manera se hace factible la participación directa (siquiera como fuente de información) de amplios colectivos de implicados que, de otra manera, debieran ser *representados*. De alguna manera, entramos de lleno en un enfoque plural.

Ahora bien, la elección de un determinado enfoque de evaluación debe estar supe- ditada fundamentalmente a los objetivos o funciones que desea cumplir. El trabajo de De Miguel (1999) introduce adecuadamente esta cuestión formulando la tensión entre dos funciones necesarias: generar conocimiento y lograr el cambio social. A estas dos funciones añade las de rendición de cuentas y mejora. Analizaremos a continuación estas funciones:

- a. La rendición de cuentas. Los actores principales de un programa o servicio no son los únicos *propietarios* del mismo. Todo programa puede identificar numerosos tipos de audiencias y, en definitiva, es la sociedad en general la *propietaria* última de cualquier intervención. De esta manera, la evaluación deberá cumplir también una función de rendición de cuentas sobre el desarrollo y resultados de un programa. Esta rendición de cuentas contará como destinatarios a los directamente afectados por el programa, a los gestores públicos y a la sociedad en general.

En las democracias representativas el debate político sobre el éxito/fracaso de los programas educativos y sociales pervierte de tal manera un análisis riguroso de la realidad de los resultados obtenidos que se hace preciso un tipo de arbitraje. La evaluación del impacto, pretendidamente neutra, debe jugar este importante papel. Los grupos comprometidos con el éxito del programa tienden a arrogarse los éxitos y externalizar los fracasos. Sus adversarios utilizan un mecanismo contrario, imputan los éxitos a causas externas y los fracasos al propio programa.

La falacia «post hoc ergo propter hoc» es utilizada sin tapujos en este debate. La visión técnica y científica debe clarificar el debate en lo posible y deslindar los efectos atribuibles al programa de aquellos que no lo son (Vedung, 1997).

- b. Fuente de conocimiento. La evaluación debe cumplir también una función de descubrimiento y desarrollo de la teoría que fundamente las intervenciones y la ciencia en general (Shadish et al. 1991).
- c. La mejora y el cambio social. La evaluación debe también ejercer un compromiso con la mejora de las condiciones de vida de la comunidad. Bajo esta perspectiva el evaluador debe procurar poner en marcha dinámicas que comprometan a las diferentes audiencias en un papel activo. Evaluación y participación de los implicados serían un binomio indisoluble.

Como puede verse, esta diversidad de funcionalidades puede generar confusión y la percepción de que son mutuamente excluyentes. Sin embargo, queremos afirmar que estas funcionalidades son complementarias y, como consecuencia, los modelos y diseños de evaluación a emplear en cada caso deberán ser *pluralistas*.

Desde esta perspectiva los modelos de evaluación debieran tener un enfoque participativo de forma que los directamente implicados cobren protagonismo en la valoración y en la toma de decisiones. Sin embargo, estas evaluaciones deben contener también importantes elementos de la mejor tradición de los métodos de investigación del área. En definitiva, las evaluaciones deberán tener una gran fundamentación en los datos, en su contenido empírico (Shadish et al. 1991). El juicio de los participantes o de los afectados por un programa no podría constituir el único criterio de mérito en la valoración. La evaluación del impacto de un programa debiera constituir una actividad rigurosa y planificada que permitiera compilar evidencias en las cuales apoyar los criterios de mérito. Esta orientación pluralista es la que puede garantizar el cumplir con las diferentes funcionalidades enumeradas anteriormente.

#### 4. DISEÑOS PARA LA EVALUACIÓN DEL IMPACTO

Bajo una orientación tradicional, los diseños para la evaluación del impacto tratan de deslindar el efecto neto de un programa de aquello que es atribuible a otros factores o circunstancias que lo acompañan. Por lo tanto, estamos plenamente en el terreno de la experimentación científica en el área de Ciencias Sociales y Humanas. Los diseños y estrategias elaborados en estos campos son los que se emplean en la evaluación del impacto. Sin embargo, la evaluación de programas ha sabido introducir también diseños de tipo opinático en los cuales expertos o implicados formulan juicios sobre los programas e intervenciones.

En el fondo de toda evaluación del impacto subyace el principio de la comparación entre el grupo que ha sido objeto de la intervención y algún grupo equivalente que no lo ha sido. Aquí reside precisamente la gran dificultad de los diseños para la evaluación del impacto: encontrar grupos de referencia realmente equivalentes con lo cuales comparar el grupo objeto de intervención. Los diseños proporcionan diversas estrategias para identificar los efectos debidos a la intervención de todo un conjunto de factores extraños o intervinientes.

Tomando como referencia el esquema propuesto por Rossi y Freeman (1993), en primer lugar, debemos distinguir entre los diseños apropiados para programas de *cobertura total* y aquellos apropiados a los programas que sólo afectan a una parte de la población denominados de *cobertura parcial*. Los programas de *cobertura total*, impiden tomar como referencia grupos que no hayan sido sometidos a la intervención y cuyas características sean similares a aquellos que sí lo han sido. En definitiva, impiden la comparación con grupos de control. La comparación deberá realizarse con el propio grupo en lo que se denominan controles *reflexivos* dentro de los cuales el diseño pretest-postest es la modalidad más conocida aunque menos recomendable.

A continuación presentaremos algunos de los principales diseños para programas de cobertura total y parcial. Debemos aclarar que los diseños que se presentan dentro del apartado de «Cobertura parcial» no pueden ser aplicados en programas de «Cobertura Total». Sin embargo, los diseños de «Cobertura total» sí pueden ser de aplicación en programas de «Cobertura parcial» aunque, en la mayoría de los casos, no son aconsejables.

#### 4.1. Algunos diseños para Programas de Cobertura Parcial

Desde el punto de vista de la validez interna, los diseños experimentales con aleatorización a todos los niveles serían la solución ideal para una correcta evaluación del impacto. Sin embargo, este tipo de diseños presentan, al menos, dos tipos de dificultades. Por un lado, en muchas ocasiones no son operativamente viables. Por otro lado, dado el nivel de control requerido, la generalización de los resultados a situaciones *naturales* puede ser cuestionable.

Dentro de los diseños experimentales con aleatorización, existen numerosas modalidades dependiendo del número de grupos que intervienen, el tipo de aleatorización, el número de medidas tomadas a través del tiempo, etc.

Los diseños experimentales garantizan la comparabilidad/equivalencia entre los grupos experimental y control mediante la aleatorización. Cuando ésta no es viable, podemos hacer uso de los denominados diseños cuasi-experimentales que intentan emular dicha comparabilidad mediante el emparejamiento u otras alternativas. Estos diseños, para ser realmente efectivos, deberán partir de un conocimiento profundo del área objeto de intervención y, por lo tanto, la especificación del conjunto de variables o factores que pudiendo interferir en los resultados del programa deberán ser controlados.

Existen diversos diseños cuasi-experimentales que se adaptan a las diferentes situaciones y necesidades de la evaluación. Los más relevantes podrían ser los siguientes (Rossi y Freeman 1993, Alvira 1991, Dendaluce 1994, Vedung 1997, Anguera 1995):

- Diseño de discontinuidad en la regresión. Estos diseños se acercan notablemente a las garantías ofrecidas por los diseños experimentales aleatorios a la hora de estimar los efectos netos reales de las intervenciones. En estos diseños la asignación al grupo objeto de tratamiento o al grupo control se realiza a partir de una puntuación de corte obtenida en una medida pretest. Son diseños con medidas pretest-posttest que permiten comprobar si las diferencias pre-post del grupo objeto del tratamiento son mayores, similares o menores que el grupo de control o de referencia.
- Diseños con grupos de comparación contruidos por emparejamiento. El grupo de referencia con el cual comparar los resultados obtenidos con el grupo objeto de intervención se selecciona en función de que tenga características similares a este último en las variables clave.
- Controles estadísticos como garantía de comparabilidad. En el diseño anterior, la comparabilidad se garantiza mediante el emparejamiento. En estos diseños se garantiza estadísticamente utilizando técnicas de tipo multivariado que neutralizan o controlan las variables en las cuales podrían diferir el grupo control y el grupo destinatario del programa. La técnicas utilizadas suelen ser el análisis de varianza o de covarianza, la regresión múltiple, etc.
- Controles genéricos. En ocasiones resulta viable disponer de datos estadísticos sobre las variables objeto de intervención en la población general. De este modo, resulta plausible la comparabilidad de los resultados obtenidos con el grupo objeto de tratamiento y los datos normativos de la población general que no ha

sido objeto de tratamiento. Las dificultades de estos diseños radican en la disponibilidad de datos sobre la población en general y en que los grupos objeto de intervención suelen ser seleccionados precisamente por tener características distintas a la población general. Sin embargo, podemos destacar por su viabilidad en el campo educativo, los diseños de *línea base no-causal construida* en sus versiones de *regresión-extrapolación* y de *comparación baremada* (Peracchio y Cook, 1987).

#### 4.2. Algunos diseños para Programas de Cobertura Total

En este tipo de programas resulta inviable obtener grupos de comparación que no hayan sido objeto de intervención. Por lo tanto, la comparación debe hacerse con el propio grupo objeto de intervención realizando comparaciones pretest-postest. A este tipo de diseños se les denomina *reflexivos* o de *controles reflexivos*.

Entre los más utilizados podemos destacar los siguientes (Rossi y Freeman 1993, Alvira 1991, Dendaluce 1994, Vedung 1997, Anguera 1995):

- Diseños simples pretest-postest. Son quizás los más conocidos y constituyen una alternativa viable cuando pueden ser obtenidas medidas de las variables relevantes con carácter previo a la intervención. Sin embargo, desde el punto de vista de la validez interna, aporta pocas garantías. La diferencia entre las medidas pre y post pueden ser debidas a factores ajenos a la propia intervención y este tipo de diseños no permite deslindar el efecto neto del programa.
- Diseños transversales para programas no-uniformes. Numerosos programas presentan diversas modalidades de aplicación por lo que, en determinada medida, podríamos hablar de varios programas distintos. Los diseños de tipo transversal son pertinentes en estas situaciones y permiten comparar los resultados de grupos sometidos a diferentes modalidades o niveles de tratamiento.
- Diseños de Panel. En estos diseños se realizan medidas sucesivas de las variables relevantes antes, durante y después de la intervención. Este análisis longitudinal permite discriminar con mayores garantías los efectos netos reales de la intervención frente a otros factores explicativos. Este tipo de diseños son también de aplicación cuando se intenta evaluar programas de aplicación no-uniforme.
- Diseños de series temporales. Este tipo de diseños son de aplicación cuando podemos disponer de un número importante de medidas de las variables relevantes tanto previas a la intervención como durante la intervención y posteriormente a la misma. El investigador comprobará si las tendencias observadas a través del tiempo mantienen su continuidad o cambian con claridad con motivo de la intervención. En definitiva, se comparan las series pre y post tratamiento.

#### 4.3. Otros diseños opináticos y la evaluación participativa

En numerosas ocasiones los diseños enumerados anteriormente no son viables por razones éticas, económicas, temporales, etc. Por otro lado, en determinadas áreas de

intervención resulta fundamental tomar como criterio de mérito o valor de un programa la opinión o satisfacción respecto al mismo de sus destinatarios, de los gestores del programa, de expertos en el área, etc. Por otro lado, si nos situamos en modelos de evaluación con enfoque participativo, las opiniones de todos las diversas audiencias pueden convertirse en elemento fundamental en la evaluación de una intervención. Por estos motivos, frecuentemente se recurre a consultar a diversas audiencias sobre su percepción del impacto real de un programa. Es lo que se denominan «shadow controls» o controles a la sombra (Rossi y Freeman 1993, Vedung 1997). En estos manuales tradicionales suelen distinguir entre las evaluaciones con base en los juicios de expertos, las que utilizan los juicios de los gestores del programa y las que toman en consideración los juicios de los destinatarios del programa. Estas alternativas no son excluyentes sino complementarias usándose con frecuencia una combinación de ellas.

Así pues, en la operativización de una evaluación que tome en consideraciones las opiniones de diversos agentes, nos podemos encontrar con una gran variedad de soluciones. Podemos destacar el modelo sueco de participante denominado *comisiones políticas ad hoc* (Vedung 1997). Este modelo intenta dar cumplimiento al mandato de participación ciudadana en el debate sobre la gestión de las políticas públicas. Según el modelo, la evaluación se lleva a cabo en cinco procesos: trabajo de la comisión, consulta, debate sobre política pública, decisión del gobierno, decisión del parlamento. En este procedimiento, el gobierno nombra la comisión y especifica los objetivos y alcance de su labor. Estas comisiones están formadas por miembros políticos, por expertos y por un secretariado. De sus debates surge un informe que es ampliamente difundido y sometido a análisis y valoración por parte de las instituciones y asociaciones afectadas.

Esta modalidad tiene un gran valor como referente de la posible operativización de una evaluación basada en opiniones y a la búsqueda del consenso en el más alto nivel de la planificación y decisión política.

Desde un abordaje menos político y más próximo, podemos rescatar aquí viejas técnicas de monitorización de equipos de trabajo que tienen una gran utilidad en los enfoques de evaluación participativa. La obra de Pineault y Davaluy (1989, p. 172) nos realiza una apropiada descripción y clasificación de estas técnicas que puede verse en la Tabla 2.

TABLA 2  
TÉCNICAS PARA ALCANZAR EL CONSENSO

Técnicas que utilizan un tipo de encuesta	. Enfoque a través de informadores clave . La técnica Delphi
Técnicas que utilizan la reflexión individual	. La técnica del grupo nominal . El brainwriting
Técnicas que utilizan la interacción	. El brainstorming . El forum comunitario
Técnicas combinadas	. Las impresiones de la comunidad

Estas técnicas sirven tanto para identificar problemas y necesidades como para el análisis y evaluación del funcionamiento de un programa y, en definitiva, de su impacto. Algunas de estas técnicas se basan en la reflexión individual (grupo nominal, *brainwriting*). Con ello se intenta evitar la superficialidad en la argumentación, los estereotipos y el parloteo. Otras técnicas intentan aprovechar la enriquecedora dinámica de la interacción en el grupo (*brainstorming*, fórum comunitario). En estas técnicas es muy importante una adecuada dinamización del grupo por parte de un especialista. Otras técnicas utilizan la encuesta como herramienta (informadores clave, técnica Delphi).

La técnica a través de *informadores clave*, no trata de alcanzar el consenso entre estos debido a que se basa en entrevistas individuales en las cuales no hay contacto o comunicación con otros informantes. El resultado de tales entrevistas podrá constatar el mayor o menor acuerdo entre los informantes.

La técnica *Delphi* sí es una técnica de grupo que puede utilizarse para alcanzar el consenso si bien la interacción entre los miembros del grupo es indirecta. Se trata de un proceso en espiral en el cual los integrantes del grupo van contestando cuestionarios que son formulados en función de las respuestas recogidas en la fase anterior hasta alcanzar el consenso o hacerse estables las divergencias. La interacción indirecta entre los informantes que asegura el anonimato de éstos es una de las grandes ventajas de este método.

La técnica del *grupo nominal*, combina una primera fase de reflexión y análisis individual (apropiada para la generación de ideas y dimensiones del problema) con una segunda fase de reflexión colectiva y una tercera de votación individual. Así pues, combina procedimientos distintos y pertinentes para cada fase de desarrollo del trabajo. La técnica del *brainwriting* se distingue de la anterior en que la formulación de las ideas iniciales la realiza el propio evaluador o monitor del grupo. La técnica del *forum comunitario* viene a ser lo que entendemos como audiencia pública.

Finalmente, el método denominado *las impresiones de la comunidad*, permite integrar y contrastar la información oficial con la información de los participantes. Combina las técnicas expuestas anteriormente en la siguiente secuencia: entrevistas con informadores clave, recogida de indicadores y datos, fórum comunitario.

#### **4.4. Enfoques integradores en la evaluación del impacto**

No puede afirmarse la superioridad de ninguna de los diseños y enfoques expuestos hasta ahora respecto a los demás. Las características concretas del programa o intervención a evaluar y aspectos de tipo económico, político, ético o temporal son los que pueden definir, en cada caso concreto, los diseños más apropiados a utilizar.

En cualquier caso, es preciso subrayar la superioridad de aquellas evaluaciones que utilicen de forma integradora diseños o enfoques diversos. En definitiva, si polarizamos la evaluación en un único tipo de diseño, aparecen sus debilidades. La diversidad de funcionalidades que debe cumplir toda evaluación hace necesaria la diversidad metodológica. Los enfoques participativos pueden ser más apropiados para una funcionalidad de compromiso con el cambio y la mejora. Sin embargo, las funcionalidades

dades de rendición de cuentas y desarrollo del conocimiento pueden requerir el uso complementario de enfoques más cuantitativos tales como indicadores de rendimiento, diseños cuasi-experimentales, etc.

Por otro lado, la triangulación que puede significar la integración de resultados obtenidos con diseños de tipo cuasi-experimental y diseños de tipo participativo u opinático es lo que puede dar una mayor solidez a las conclusiones finales de una evaluación. De alguna manera, venimos a repetir aquí los tan citados argumentos manejados por Cook y Reichardt (1986) respecto a la complementariedad de los métodos cualitativos y cuantitativos en la investigación evaluativa. Apostamos por lo tanto por la complementariedad e integración de enfoques diversos en lo que puede denominarse evaluación pluralista, multimétodo. Según recoge apropiadamente De Miguel en esta misma obra, la investigación social en general y la investigación evaluativa en particular, destacan por ser áreas donde esta complementariedad está más extendida y las discusiones paradigmáticas más superadas. Las diversas funciones que ha de cumplir la evaluación (rendición de cuentas, compromiso con el cambio social y desarrollo del conocimiento), así lo demandan.

## **5. PRODUCTIVIDAD Y EFICIENCIA: LA EVALUACIÓN ECONÓMICA**

La valoración del éxito de un programa de intervención debe considerar también los recursos utilizados. La continuidad o generalización de los programas suele depender en gran medida de la rentabilidad constatable de los recursos o esfuerzos requeridos.

La dimensión de productividad simplemente divide un indicador de los resultados de un determinado programa o servicio entre los recursos utilizados. Esos recursos pueden estar medidos en términos monetarios, en horas trabajadas, locales utilizados, etc. Las ratios resultantes resultan difíciles de valorar en términos absolutos por lo que es recomendable tomar como referencia medidas de productividad en ejercicios anteriores, productividad de programas o servicios similares, etc.

La aparente simplicidad operativa del cálculo de la productividad esconde grandes dificultades en la operativización de las medidas de resultados y de costes. Por un lado, los programas o intervenciones suelen dar lugar a resultados múltiples. Por otro lado, no todos los resultados resultan sencillos de cuantificar. En definitiva, la elaboración de un índice de resultados obtenidos suele ser parcial y cuestionable. Por otro lado, determinar los costes en términos monetarios o en cualquier otra unidad resulta también especialmente complicado. Los costes pueden ser directos pero también indirectos. Los programas y servicios desarrollan una diversidad de funciones siendo difícil la imputación de recursos consumidos a cada una de dichas funciones o propósitos. Además, la productividad puede olvidar con facilidad el criterio de calidad/ utilidad de los resultados o servicios prescindiendo de los fines últimos de éstos.

La evaluación económica de un programa puede ser más o menos completa. Una buena clasificación de los diversos tipos o niveles de evaluación económica podemos encontrarla en obras tales como: Drummond, Stoddart y Torrance (1991). Para estos autores, una evaluación económica completa debiera comparar dos o más alternativas

de intervención y examinar tanto los costes como los resultados. La no consideración de alguno de estos elementos o de ambos da lugar a diversos tipos de evaluación parcial. La Tabla 3 adjunta, tomada de la citada obra, clarifica esta clasificación.

TABLA 3  
LA EVALUACIÓN ECONÓMICA SEGÚN DRUMMOND,  
STODDART Y TORRANCE (1991)

¿Se examinan tanto los costes (entradas) como los resultados (salidas) de las alternativas?

		NO		SÍ
¿Hay comprobación entre dos o más alternativas?	NO	Se examinan sólo los resultados	Se examinan sólo los costes	Descripción del coste-resultado
		<i>Descripción del resultado</i>	<i>Descripción del coste</i>	
	SÍ	<i>Evaluación de la eficacia o la efectividad</i>	Análisis del coste	EVALUACIÓN ECONÓMICA COMPLETA <i>Anál. coste-minimización</i> <b>Análisis coste-efectividad</b> <i>Análisis coste-utilidad</i> <b>Análisis coste-beneficio</b>

Como puede observarse sólo podemos considerar como evaluaciones económicas completas aquellas que toman en consideración tanto los costes como los resultados además de comparar entre dos o más alternativas. A continuación describiremos brevemente tres de los tipos de evaluación económica más frecuentes.

En la dimensión de eficiencia podemos distinguir entre el coste-beneficio y el coste-efectividad. El coste-beneficio divide el valor (en términos monetarios) de los efectos de un programa entre el coste (también en términos monetarios). El coste-efectividad divide un indicador de los efectos (sustantivos en relación a los objetivos) de un programa en unidades físicas entre los costes en términos monetarios.

La estimación de los beneficios de una intervención presenta especiales dificultades en los programas sociales donde resulta difícil asignar un valor monetario a sus resultados. En general, resulta más viable asignar un valor monetario a los inputs de un programa que a sus resultados. Esta es una de las razones que apoyan el uso de los análisis coste-efectividad en lugar de los análisis coste-beneficio. El coste-efectividad cuantifica en términos monetarios solamente los inputs o recursos utilizados por el programa mientras que, en el denominador de la función, emplea una cuantificación

de los beneficios en unidades físicas o de otro tipo. Estos beneficios deben referirse a elementos sustantivos de los objetivos del programa.

No podemos finalizar sin hacer mención a la evaluación del coste-utilidad. Este tipo de evaluación es cada vez más utilizada en todos aquellos programas o intervenciones de carácter público que tengan que ver con la prestación de algún servicio directo o indirecto al ciudadano. En definitiva, se trata de comparar los costes con su utilidad operativizada a través de las percepciones que los destinatarios tienen del programa-intervención... Con ello, la evaluación económica enlaza también con los modelos de evaluación participativa donde la percepción e interpretación de los destinatarios del servicio se convierte en elemento esencial de la misma.

## REFERENCIAS

- ALVIRA, F. (1991). *Metodología de la evaluación de programas*. Madrid: C.I.S.
- ANDER-EGG, E. (1990). *Repensando la Investigación-Acción-Participativa*. Vitoria-Gasteiz: Servicio de Publicaciones del Gobierno Vasco.
- ANGUERA, M.T. (1995). Diseños. En R. Fernández-Ballesteros (Ed.): *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid: Síntesis.
- COOK, T.D. y REICHARDT, Ch. S. (1986). *Métodos cuantitativos y cualitativos en investigación evaluativa*. Barcelona: Morata.
- DE MIGUEL, M. (1995). Participación ciudadana en la evaluación de las políticas públicas. *Aula Abierta*, 66, 211-218.
- DE MIGUEL, M. (1998). *Modelos y diseños en la evaluación de programas*. Documento no publicado.
- DE MIGUEL, M. (1999). *La evaluación de programas: entre el conocimiento y el compromiso*. IX Congreso nacional de modelos de investigación educativa. Málaga 21-23 de Octubre.
- DENDALUZE, I. (1994). Diseños cuasiexperimentales. In V. GARCÍA HOZ (Eds.), *Problemas y métodos de investigación en educación personalizada* (pp. 286-304). Madrid: Rialp.
- DRUMMOND, M.F., STODDART, G.L. y TORRANCE, G.W. (1991). *Métodos para la evaluación económica de los programas de atención de la salud*. Madrid: Díaz de Santos.
- FETTERMAN, D.M., KAFTARIAN, S.J. y WANDERSMAN, A. (1996). *Empowerment evaluation: Knowledge and tools for self-assessment & accountability*. Thousand Oaks CA: Sage.
- FLOC'HLAY y PLOTTU, E. (1998). Democratic Evaluation. *Evaluation*, 4, 3, 261-277.
- GUBA, E.G., & LINCOLN, Y.S. (1981). *Effective evaluation*. San Francisco: Jossey Bass.
- GUBA, E.G., & LINCOLN, Y.S. (1989). *Fourth generation evaluation*. Newbury Park, California: Sage.
- PINEAULT, R. y DAVELUY, C. (1989). *La planificación sanitaria: conceptos, método, estrategias*. Barcelona: Masson.
- PERACCHIO, L.A. y COOK, T.D. (1988). Avances en el diseño cuasi-experimental. En I. DENDALUZE (Ed.): *Aspectos metodológicos de la investigación educativa*. Madrid: Narcea.

- REBIEN, C.C. (1997). Development assistance evaluation and the foundations of program evaluation. *Evaluation Review*, 21, 4, 438-460.
- ROSSI, P.H., & FREEMAN, H.E. (1989). *Evaluación. Un enfoque sistemático para programas sociales*. México, D.F.: Trillas.
- ROSSI, P.H., & FREEMAN, H.E. (1993). *Evaluation: A systematic approach*. Newbury Oark: Sage.
- SCHALOCK, R.L. (1995). *Outcome-Based Evaluation*. Plenum Press: New York.
- SCRIVEN, M. (1991). Beyond formative and summative evaluation. In M. W. McLAUGHLIN & C. PHILLIPS (Eds.), *Evaluación and Education: At quarter Century* (pp. 19-64). Chicago: NSSE.
- SCRIVEN, M.S. (1983). Evaluation Ideologies. In G.F. MADAUS, M.S. SCRIVEN, & D.L. STUFFLEBEAM (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* Boston: Kluwer.
- SHADISH, W.R., COOK, T.D. & LEVITON, L.C. (Ed.). (1991). *Foundations of program evaluation: theories of practice*. Newbury Park, California: Sage.
- STAKE, R.E. (1983). Program evaluation, particularly responsive evaluation. In G.F. MADAUS, M.S. SCRIVEN & D.L. STUFFLEBEAM (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* Boston: Kluwer.
- STAKE, R.E. (1991). Retrospective on «The countenance of educational evaluation». In M.W. McLAUGHLIN & C. PHILLIPS (Eds.), *Evaluación and Education: At quarter Century* (pp. 67-88). Chicago: NSSE.
- STUFFLEBEAM, D.L. & SHINKFIELD, A.J. (1993). *Evaluación sistemática: guía teórica y práctica*. Barcelona: Paidós-MEC.
- TYLER, R.W. (1991). General stament on program evaluation. In M.W. McLAUGHLIN & C. PHILLIPS (Eds.), *Evaluación and Education: At quarter Century* (pp. 3-17). Chicago: NSSE.
- TYLER, R.W., MADAUS, G.F. & STUFFLEBEAM, D.L. (Ed.). (1989). *Educational evaluation: classic works of Ralph W. Tyler*. Boston: Kluwer.
- VEDUNG, E. (1997). *Evaluación de políticas públicas y programas*. Madrid: Instituto Nacional de Servicios Sociales.