



Aspectos teóricos de la evaluación de las políticas públicas

Talina Merit Olvera Mejía

UAEH
Universidad Autónoma del Estado de Hidalgo

Aspectos teóricos de la evaluación de las políticas públicas

Instituto de Ciencias Sociales y Humanidades
Área Académica de Ciencias Políticas y Administración Pública



CONSEJO
EDITORIAL

La publicación de este libro se financió con recursos del Programa de Fortalecimiento de la Calidad Educativa (*PFCE*) 2017.

Aspectos teóricos de la evaluación de las políticas públicas

Talina Merit Olvera Mejía



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
Pachuca de Soto, Hidalgo, México
2023

UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO

Octavio Castillo Acosta
Rector

Julio César Leines Medécigo
Secretario General

Marco Antonio Alfaro Morales
Coordinador de la División de Extensión de la Cultura

Ivonne Juárez Ramírez
Directora del Instituto de Ciencias Sociales y Humanidades

Fondo Editorial

Asael Ortiz Lazcano
Director de Ediciones y Publicaciones

Joselito Medina Marín
Subdirector de Ediciones y Publicaciones

Primera edición electrónica: 2023

D.R. © UNIVERSIDAD AUTÓNOMA DEL ESTADO DE HIDALGO
Abasolo 600, Col. Centro, Pachuca de Soto, Hidalgo, México, C.P. 42000

Dirección electrónica: editor@uaeh.edu.mx

El contenido y el tratamiento de los trabajos que componen este libro son responsabilidad de los autores y no reflejan necesariamente el punto de vista de la Universidad Autónoma del Estado de Hidalgo.

ISBN: 978-607-482-767-5

Esta obra está autorizada bajo la licencia internacional Creative Commons Reconocimiento – No Comercial – Sin Obra Derivada (by-nc-nd) No se permite un uso comercial de la obra original ni la generación de obras derivadas. Para ver una copia de la licencia, visite <https://creativecommons.org/licenses/by-nc-sa/4.0/>.



Hecho en México/Printed in México

Contenido

Introducción	9
<hr/>	
1. El análisis de políticas públicas. Evaluación dentro del ciclo de políticas públicas	13
<hr/>	
2. Bases conceptuales de la evaluación de programas y políticas públicas	19
2.1 Concepto de evaluación	19
2.2 Evaluación de programas y políticas públicas. Hacia una visión amplia e integral de la evaluación	21
2.3 Funciones de la evaluación. ¿Para qué nos sirve evaluar?	23
2.4 Propuestas para una mayor utilización de las evaluaciones	28
<hr/>	
3. El innegable componente político en la evaluación. Actores y contexto	35
<hr/>	
4. La evaluación a través de los años. De la evaluación como medición de resultados a la evaluación como aprendizaje organizacional	41
<hr/>	
5. Los métodos cuantitativos y cualitativos. La importancia de los paradigmas en la evaluación	49
5.1 Método mixto. La combinación de métodos cualitativos y cuantitativos en la evaluación	55
<hr/>	

6. Los enfoques evaluativos más relevantes para inicios del presente siglo	59
6.1 Evaluación por objetivos	65
6.2 Modelos experimentalistas	65
6.3 Estudios orientados a la decisión y a la rendición de cuentas	67
6.4 Estudios orientados al consumidor	70
6.5 Enfoques de acreditación y certificación	71
6.6 Evaluación centrada en la utilización	72
6.7 Estudios centrados en el cliente (o evaluación sensible)	74
6.8 Evaluación democrática deliberativa	76
6.9 Evaluación constructivista	77
6.10 Evaluación de estudio de caso	79
6.11 Evaluación de resultados como valor añadido	80
<hr/>	
7. Tipos de evaluación	83
7.1 La evaluación según la función que cumple	83
7.2 La evaluación según el momento en que se realiza	84
7.3 La evaluación según los agentes que la realizan	85
7.4 La evaluación de acuerdo con las fases del programa	85
7.5 Otros tipos de evaluaciones	87

8. El proceso de evaluación	91
8.1 Lista de control para diseñar evaluaciones de Daniel L. Stufflebeam (2007)	92
8.2 Planteamiento de la evaluación. Evaluabilidad del programa	97
8.3 Establecimiento de los criterios de valor	98
8.4 Selección de las operaciones a observar	100
8.5 Selección del diseño de evaluación	102
8.6 Recogida de información	103
8.7 Análisis de datos: enjuiciamiento y recomendaciones	104
8.8 Informe	105
8.9 Administración de la evaluación	107

9. La evaluación de diseño y gestión de programas	109
9.1 La evaluación del diseño de programas. Concepto y objetivo	109
9.2 Criterios para evaluar el diseño del programa	110
9.3 La evaluación del sistema de gestión e implementación de programas. Concepto y objetivo	120
9.4 Criterios para evaluar la gestión del programa	123

Conclusiones	129
Referencias	133

Introducción

El gobierno mexicano comienza a aceptar que la evaluación es una realidad que no puede evadir si quiere mejorar las acciones gubernamentales y lograr los resultados que se ha planteado. Reconoce también que llevar a cabo evaluaciones le otorga al mismo tiempo legitimidad, pues sirve como mecanismo de rendición de cuentas y apoyo para tomar decisiones de manera más informada. En México, la evaluación ha sido una herramienta que tradicionalmente no ha sido aprovechada ni usada por los gobiernos en la mejora de las políticas y los programas públicos. El limitado desarrollo teórico de la evaluación dentro del análisis y ciclo de políticas públicas, se ha visto superado por la práctica y obligatoriedad que ha adquirido la evaluación dentro de todos los ámbitos del gobierno. En México, académicamente ha habido poca referencia al tema, solo en años recientes a raíz de la creación de instituciones dedicadas a la evaluación de programas es que comienzan a aparecer publicaciones¹ o seminarios.

La aprobación y reforma de diversos documentos normativos durante este siglo, han obligado a que la práctica evaluativa sea más frecuente. En el año 2000 el Congreso solicitó de manera obligatoria la evaluación a los programas federales de desarrollo social sujetos a reglas de operación, a

1 Cardozo (2015) realiza un análisis respecto de la producción de publicaciones sobre evaluación en México. Establece que en los últimos 10 años se dispone alrededor de 15 publicaciones, entre ellas menciona las de Mejía (2003), Magaña y Vargas (2004), Cardozo (2006), Vargas (2007), Castro (2007), Cortés *et al.* (2008), Cardozo (2009), Ramos *et al.* (2011), Cardozo (2012a y b), Cardozo y Mundo (2012), Arellano (2012), Ramos (2012), Ramos (2013) y Maldonado y Galíndez (2013).

partir de entonces se dieron pasos para institucionalizar la evaluación en México.

En 2004, con la aprobación de la Ley General de Desarrollo Social se dedicó un apartado exclusivo a la Evaluación de la Política de Desarrollo Social (artículos del 72 al 85), se establecía que esta evaluación estaría a cargo del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval). Dos años después, en 2006, se creó dicho consejo. En ese mismo año, el Congreso aprobó la Ley Federal de Presupuesto y Responsabilidad Hacendaria, que crearía el Sistema de Evaluación de Desempeño, impulsado a partir de 2007 junto con el Presupuesto Basado en Resultados por la Secretaría de Hacienda y Crédito Público (SHCP).

En 2007, la SHCP, la Secretaría de la Función Pública (SFP) y el Coneval, publicaron los Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal, con el objetivo de regular la evaluación de los programas federales, la elaboración de la matriz de indicadores y los sistemas de monitoreo. Además, en 2014 se adiciona el apartado C al Artículo 26 de la Constitución Política de los Estados Unidos Mexicanos, que establece que el Estado contará con un Consejo Nacional de Evaluación de la Política de Desarrollo Social a cargo de la medición de la pobreza y de la evaluación de los programas y acciones de dicha política. Todos estos elementos constituyen la Política Federal de Evaluación de México.

La escasez de trabajos académicos sobre evaluación de políticas públicas y de programas, además de su mayor impulso y obligatoriedad en todos los niveles de gobierno, han motivado el conocer más acerca de esta fase de las políticas públicas, relegada a la etapa final de dicho ciclo, encargada tradicionalmente de enfocarse en los resultados. Este mayor conocimiento de la evaluación hace imprescindible su estudio de una manera clara y sistemática, de modo que pueda contribuir al análisis y práctica de la misma.

El objetivo de este libro es presentar al lector aspectos básicos de la evaluación para un acercamiento y conocimiento de la misma. El propósito es servir como una guía básica a quienes busquen aproximarse al tema de la evaluación. No pretende ser exhaustivo ni profundizar en cada uno de estos aspectos, sino más bien acercar al interesado en los temas relevantes y actuales en el ámbito de la evaluación de políticas públicas y programas, y de los cuales es preciso conocer para ejercer y mejorar esta práctica. Los

temas se abordan a través de la revisión de la literatura, donde se incluyen las principales aportaciones de los referentes en la evaluación a nivel internacional (Estados Unidos y Francia) y en lengua española (España y México). La bibliografía presentada al final del libro sirve de referencia para acercarse y profundizar en cada uno de los temas presentados.

La estructura del libro comienza con un apartado dedicado al análisis de políticas públicas y esa doble dimensión descriptiva y prescriptiva, propuesta por Laswell, el propósito es entenderlo y ubicar a la evaluación dentro de dicho análisis (análisis marcado en un inicio por el interés exclusivo en la toma de decisiones). Se discute el papel que juega la evaluación dentro de este análisis, haciendo referencia a la etapa final que ha desempeñado tradicionalmente. Sin embargo, se señala la importancia de ver a la evaluación como un proceso integral dentro de las políticas públicas y programas, aspecto que varios autores han desarrollado y cuyo trabajo se analiza.

En otro apartado se discute la definición de evaluación, se hace mención a las diferencias y/o similitudes entre evaluar políticas y programas. Otro tema relevante son las funciones básicas que debe cumplir la evaluación; sin embargo, se plantean también aquellas razones menos legítimas al momento de solicitarla y realizarla. De relevancia actual y con la intención de que la evaluación tenga un uso real y, entre otras cosas, apoye la toma de decisiones, el tema de su uso se aborda en este libro. Se señalan algunas razones por las cuales no se atiende a sus resultados y, basado en aportes de diversos estudiosos del campo de la evaluación, se exponen algunas propuestas para fomentar su uso.

Durante el libro se insiste además en que la evaluación no es solo una tarea técnica y metodológica sino una tarea de relaciones interpersonales que se dan en un contexto político, por lo que se analiza a los actores y contexto de la evaluación, resaltando el componente político en las evaluaciones.

Para entender el avance en la práctica de la evaluación, se analiza su evolución a través de los años y las condiciones que han influido en cada una de sus etapas. Iniciamos con los aportes realizados en Estados Unidos por parte de Tyler y Campbell y Stanley, y su evaluación por objetivos y experimental, respectivamente, hasta llegar a los aportes más recientes que tienen que ver con la evaluación como aprendizaje organizacional, pasando por el uso efectivo de la evaluación y que involucra cuestiones políticas y

de negociación. A raíz de que el debate entre los métodos cuantitativo y cualitativo ha permeado la evaluación, se examinan las posturas de cada uno de ellos para establecer cuál sería el método más adecuado en la evaluación. En el entendido de que la historia de la evaluación ha quedado marcada en diversos enfoques evaluativos propuestos por diferentes autores, se hace un repaso por los más sobresalientes, y que han marcado la historia evaluativa, hasta llegar a aquellos enfoques que muestran su validez para este siglo. Asimismo se aborda la tipología de la evaluación.

Dado que la evaluación es una actividad práctica, se hace referencia su proceso de realización: su planteamiento, su diseño, la recogida y análisis de datos, la elaboración del informe, y otros aspectos administrativos de la evaluación. Ante la relevancia del carácter valorativo de la evaluación, y del hecho de que para poder realizar esa valoración se necesita establecer antes ciertos criterios contra los cuales juzgar el objeto evaluado, se analiza el establecimiento de los criterios de valor. Finalmente, el texto aborda algunos criterios para evaluar el diseño y la gestión de los programas.

1. El análisis de las políticas públicas. Evaluación dentro del ciclo de políticas públicas

Es común que al hablar de política asociemos el término a los mecanismos de lucha de poder, a cuestiones electorales, a las acciones de los partidos políticos, a los puestos de elección popular o a los grupos de interés. Sin embargo, esto es lo que en inglés se conoce con el término *politics*. Los trabajos de Lindblom o Wildavsky han permitido ampliar el horizonte del término política, y al día de hoy política no se limita solo a lo antes mencionado. Política, o *policy* en inglés, también se refiere a lo que Meny y Thoenig (1992: 89) han llamado “un marco de orientación para la acción, un programa o una perspectiva de autoridad”. Aguilar Villanueva entre las varias definiciones que establece del término destaca el doble sentido que se le da como curso de acción, al determinar que la política que se decide y se quiere llevar a cabo es:

...por una parte, un conjunto más o menos interrelacionado de acciones, que en razón de los recursos con los que cuentan los actores, los medios que emplean y las reglas que siguen es considerado idóneo y eficaz (o eficiente) para realizar el estado de las cosas preferido... Por otra parte, el curso de acción que se desata y que efectivamente es desarrollado, intencional o reactivamente, por los varios actores gubernamentales, por los otros poderes estatales, por la multitud de diferentes actores políticos, grupos económicos y organizaciones sociales, con

sus interacciones cooperativas o conflictivas, puede distanciarse más o menos de la estrategia intencional de acción colectiva, diseñada y organizada por la autoridad estatal (Aguilar, 1996: 25).

Aguilar Villanueva concluye que una política es un “diseño de acción colectiva intencional, que abarca también el curso que efectivamente toma la acción como resultado de muchas decisiones e interacciones así como los hechos reales que la acción colectiva produce”. Entonces:

... las ‘políticas’ se ocupan de aquellas esferas consideradas como ‘públicas’ (...). La idea de las políticas públicas presupone la existencia de una esfera o ámbito de la vida que no es privada o puramente individual, sino colectiva. Lo público comprende aquella dimensión de la actividad humana que se cree que requiere la regulación o intervención gubernamental o social, o por lo menos la adopción de medidas comunes (Parsons, 2007: 37).

Tamayo Sáez (1997: 281) define a la política pública como “el conjunto de objetivos, decisiones y acciones que lleva a cabo un gobierno para solucionar los problemas que en un momento determinado los ciudadanos y el propio gobierno consideran prioritarios”. Así entonces, parece lógico que toda acción de política pública deba estar orientada a producir bienestar social y, en este sentido, podemos decir que la fundamentación de las políticas públicas, de acuerdo con Monnier (1995), reside en la definición de la demanda social, es decir, en identificar los problemas sociales que requieren la intervención de los poderes públicos. Sin embargo, una política pública no solo consiste en acciones, sino también en todo aquello que los actores gubernamentales deciden no hacer (Dye, 1972: 18 y 1984: 1), estudian “qué hacen los gobiernos, por qué lo hacen y cuál es su efecto” (Dye; 1976: 1).

Aunado a lo anterior, y considerando que los poderes públicos deben dar respuesta de una manera eficaz a las demandas de la sociedad (después de todo, los gobiernos se encaminan a resolver problemas sociales), es importante saber qué instrumentos pueden ayudar a estudiar las acciones de los poderes públicos. Saber por qué se ha elegido cierta cuestión para

actuar, por qué se optó por esa intervención, cuáles son los objetivos que se pretenden alcanzar, cuánto tiempo se debe actuar sobre el problema, quiénes serán los beneficiarios y a quiénes se afectará con la decisión, cuál es el mejor camino para poner en práctica el programa, cuáles serán los resultados de dicha intervención y finalmente saber si seguir o no con la intervención pública. El estudio y análisis de las políticas públicas responde a estas cuestiones y permite conocer las acciones y productos de quienes gobiernan. Como mencionan Meny y Thoenig (1992: 94), se trata de una tentativa metodológica para lograr un mejor dominio de la gestión de las políticas públicas, es decir, de la transformación de productos en efectos o impactos.

Inicialmente, los analistas (y el análisis) de políticas públicas se centraban únicamente en el mejoramiento de la toma de decisiones, las investigaciones tenían un enfoque práctico y que fuera útil para la acción, se limitaban a la formulación del problema y a la búsqueda de una solución, los analistas no cuestionaban las decisiones y mucho menos las acciones que seguían a esa decisión. Pero actualmente, a los analistas de políticas públicas les interesan “los ‘problemas públicos’ y su relación con las políticas públicas, el contenido de las políticas públicas, lo que los tomadores de decisiones y los diseñadores de políticas hagan o dejen de hacer, los insumos y el proceso de determinada área de las políticas públicas y las consecuencias de las políticas públicas en cuestión de los resultados y los impactos” (Parsons, 2007: 64). Es decir, se ha ampliado el campo de estudio hasta intentar responder a las preguntas que se plantean en cada una de las etapas de la política pública o programa. El análisis de políticas públicas toma como unidad de análisis a las políticas y, en consecuencia, a la acción de las instancias de gobierno.

En este sentido, las políticas públicas desarrollaron una doble dimensión donde, de acuerdo con Lasswell (1951; 1970) podían definirse en función del conocimiento en el proceso de las políticas y el conocimiento de las políticas. La primera dimensión tiene un sentido más prescriptivo y normativo. Se refiere al interés en y para el proceso de la política (*knowledge in*), el cual usa métodos de investigación, análisis de las ciencias sociales, ciencias naturales, ciencias de la decisión, así como de los métodos analíticos para desarrollar la definición de los problemas, la toma de decisiones, la ejecución y la evaluación de las políticas con el propósito de modificar y mejorar la decisión pública e incrementar la

racionalidad en todo ese proceso. La otra dimensión tiene una orientación más descriptiva y positiva, se ocupa del conocimiento del proceso de toma de decisiones (*knowledge of*) a través de estudios empíricos sobre cómo se elaboran y llevan a cabo las políticas, más específicamente “cómo se definen los problemas y las agendas, cómo se formulan las políticas públicas, cómo se toman las decisiones y cómo se evalúan e implementan las políticas públicas” (Parsons, 2007: 32). Para Laswell estas dos actividades son interdependientes y complementarias, y añade que el enfoque de políticas tiende a la contextualización, a la diversidad de los métodos utilizados y se orienta hacia los problemas.

Uno de los enfoques para explicar las políticas es el “enfoque por etapas”, el cual las concibe como un proceso secuencial, derivado de Lasswell. Este analiza las políticas públicas como procesos que empiezan con la definición de la agenda y terminan con la evaluación y el resultado de las políticas públicas (Parsons, 2007: 73). Subirats basado en los estudios de Hogwood-Gunn (1984), Thoening (1985), Ham-Hill (1984) y Jones (1984), señala que:

... el esquema de análisis comprendería la percepción y definición del problema; los intereses afectados; el grado de organización; el acceso a los canales representativos; la consecución del estatuto propio de ‘tema’ a incluir en el programa o agenda de actuación de los poderes públicos; la formulación de una solución o de una acción de respuesta; el establecimiento de objetivos y prioridades; los soportes políticos, presupuestarios y administrativos de la solución propuesta; la implementación o puesta en práctica de esa política; la evaluación y control de los efectos producidos; y el mantenimiento, revisión o terminación de esa política (Subirats, 1992: 42).

Para Tamayo Sáez (1997: 281) las políticas públicas se pueden entender como un proceso de toma de decisiones que se inicia cuando un gobierno o un director público detecta la existencia de un problema que, por su importancia, merece su atención y termina con la evaluación de los resultados que han tenido las acciones impulsadas para mitigar o corregir ese problema.

En este sentido, las políticas públicas se han dividido en etapas, en un proceso que sugiere algo ordenado y racional, pero es un hecho que las políticas públicas y programas rara vez son eso. Sin embargo, esta división sirve como una guía, y ha ayudado a hacer más sencillo el estudio y análisis de dichas políticas. Hay que tener en cuenta que ninguna de estas fases debería ser analizada en solitario, ni tampoco intentar dar respuestas a una fase sin considerar a las demás, pues cualquier variación en alguna de ellas afecta al resto de las etapas. Este ciclo de construcción de las políticas públicas consiste en: identificar y definir el problema, formular las alternativas de solución, adoptar una alternativa, implantar la alternativa seleccionada, y evaluar los resultados obtenidos. La tabla 1 muestra la relación entre el ciclo de políticas públicas y su análisis, los cuales van de la mano y no se puede concebir una fase de política sin su respectivo análisis:

Tabla 1 Relación entre el ciclo de políticas públicas y su análisis

ACCIÓN	ANÁLISIS
1.- Identificación de problemas.	1.- Diagnósis: ¿cuál es el problema?, ¿qué sucederá si no actuamos?
2.- Formulación de políticas.	2.- ¿Cómo formular el plan de acción? ¿Qué objetivos? ¿Opciones? ¿Efectos?
3.- Decisión.	3.- ¿Factores políticos presentes?
4.- Implementación.	4.- ¿Influencias en el proceso de puesta en práctica? ¿Información?
5.- Evaluación.	5.- ¿Definición de éxito? ¿Estándares de evaluación? ¿Consecuencias?

Fuente: Starling, G., 1988: 10, en Subirats, 1992: 43.

Como se puede ver, la evaluación forma parte de las políticas públicas y de su respectivo análisis, es por eso que se le ubica dentro de este marco de referencia. Algo a destacar en este ciclo es que la evaluación se

considera la parte final, la terminación de la política, y en cierta medida la retroalimentación del proceso. La idea de evaluación está vinculada a valorar los resultados e impacto de la política y apoyar en la decisión de si se continúa, modifica o termina con dicha política, realizándose así el proceso de retroalimentación. Actualmente, tanto en el plano teórico como empírico, los encargados de las políticas y programas también se hacen preguntas no solo sobre los resultados de la política o programa, sino también sobre su diseño y proceso de implementación, ampliándose el campo de estudio y actuación.

Basados en lo anterior, ha resultado común que al análisis de políticas (*policy analysis*) se le considere prospectivo, anticipatorio, “dirigido a decisiones que aún se han de tomar” (Beckman, 1977: 222 cit. por Subirats, 1994: 139), y la evaluación de programas se vea más como un enfoque retrospectivo, “para, mirando al pasado, facilitar vías de acción de futuro” (Poland, 1974: 333 cit. por Subirats, 1994: 139), realizándose a políticas o programas que ya han sido puestos en práctica. Sin embargo, como menciona María Bustelo (2001: 25), este carácter retrospectivo de la evaluación no queda tan claro a medida que se va ampliando el foco de la evaluación, además de que en algunos países se realizan evaluaciones “*ex-ante*”, lo que hace difícil su clasificación retrospectiva.

Los estudios de políticas públicas han contado a través de los años con aportaciones principalmente en las fases de toma de decisiones e implementación ², la evaluación había quedado relegada y cuando se estudiaba se le daba un carácter de conclusiva, de evaluación de resultados e impactos. Sin embargo, desde finales del siglo pasado la evaluación ha mostrado su importancia y cada vez más surge el interés por realizar tanto trabajos de manera empírica y a programas, como de estudios que analizan y valoran los procesos de evaluación. Poco a poco se ha ido considerando la evaluación como un proceso integral, donde constantemente aparecen nuevos enfoques, posturas y visiones sobre esta parte esencial de las políticas públicas. Todo ha contribuido al avance en el logro de un equilibrio entre el trabajo empírico-profesional y el académico-de investigación.

2 Los trabajos de Harold D. Lasswell, Theodore J. Lowi, Charles E. Lindblom o Giandomenico Majone, son algunos ejemplos de las aportaciones al proceso de toma de decisiones. En cuanto a la implementación están los estudios de Pressman, Widavsky, Bardach, Lipsky o Elmore.

2. Bases conceptuales de la evaluación de programas y políticas públicas

2.1 Concepto de evaluación

A diferencia de su equivalente en inglés, en español se suele utilizar el término evaluación para englobar diversas actividades que si bien son similares cada una se distingue entre sí. Por ejemplo, en inglés existen los términos *evaluation*, *appraisal* o *assessment* (valoración); *evaluation* se usa para referirse a objetos y *assessment* es usado para evaluar personas. Sin embargo, en castellano se emplea el término evaluación para englobar todos estos aspectos. Evaluación proviene del francés *évaluer* y es también un derivado del latín *valere* que significa valorar, lo cual implica la acción de fijar valor a una cosa. Los usos de la evaluación formal pueden remontarse al año 2000 a.C. cuando los oficiales chinos realizaban exámenes en el sector público de servicio civil para medir la eficiencia de los aspirantes a posiciones del gobierno (Fitzpatrick, 2011). En años más recientes, en el siglo XX, Scriven (1967) profundizó en la idea de evaluación como ciencia de la valoración, para él “su principal característica es que supone la estimación del mérito de un objeto” y establece que esta actividad no solo consiste en revisar que las cosas se hagan bien y de acuerdo con lo establecido, implica ir más allá, emitiendo un juicio de valor sobre lo observado, si no se emite un juicio de valor no hay evaluación, sino auditoría, investigación o control.

En su concepto de evaluación, Scriven menciona los términos mérito (*merit*) y valor (*worth*), que también son mencionados por otros autores,

y aunque para nosotros su significado y aplicación es prácticamente el mismo, existe una diferencia entre ambos términos. La diferencia radica en que el mérito es algo intrínseco al objeto, mientras el valor es extrínseco. Es decir, la acción o programa, en palabras de Stufflebeam y Shinkfield (1987: 21), “¿hará bien lo que se supone debe hacer? Si lo hace, será meritorio, pero puede que no sea válido” para la sociedad o contexto en el que esa acción o programa es aplicado.

El hecho de que evaluación es emitir un juicio de valor, no se discute por los estudiosos en este ámbito (Scriven, 1967; Rutman, 1980; Joint Committe, 1981, 1994; Cheimsky, 1985; Rossi y Freeman, 1985, 1993; Stufflebeam y Shinkfield, 1987; Alvira, 1992; Shadish, Cook y Levinton, 1991; Meny y Thoenig, 1992; Vedung, 1997; Weiss, 1998; Bustelo, 2001; Fernández Ballesteros, 2001; Cardozo, 2006), se le considera la parte esencial de esta actividad y sin la cual no está culminada la tarea evaluativa. Pero si se trata de emitir un juicio de valor, es necesario establecer antes unos criterios contra los cuales juzgar el objeto evaluado, siendo esta otra actividad inherente a la evaluación. Los juicios de valor deben emitirse y estar sustentados sobre la evidencia hallada, sobre los datos empíricos relativos al programa o política pública.

Los autores coinciden también en que para realizar una evaluación esta debe hacer uso de los métodos de las ciencias sociales o de la investigación social y llevarse a cabo de manera sistemática³. Se enfatiza que los juicios de valor deben basarse en una investigación sistemática y rigurosa (Joint Committe, 1981, 1994; Chelinsky, 1985; Rossi y Freeman, 1985; Patton, 1987; Alvira, 2002; Menny y Thoenig, 1992; Fink, 1993; Vedung, 1997; Weiss, 1998; Mejía Lira, 2003). Aunado a lo anterior, la evaluación necesita emitir recomendaciones sobre lo previamente valorado. Esta última actividad permitirá establecer de manera más clara las acciones que tenderán a la mejora o reorientación de la política o programa público

3 Estas dos características se refieren concretamente al sentido metodológico de la evaluación, con el propósito de guardar la objetividad y veracidad en los datos; sin embargo, actualmente el cumplir con el rigor metodológico de la evaluación no es suficiente para que una evaluación sea válida y de utilidad para quienes la han solicitado y para la audiencia afectada por la misma y por el programa. Hoy se agrega al aspecto metodológico, la participación de quienes encargan la evaluación y de todas las partes interesadas en la misma, por tanto, integrar el aspecto político inherente a cada evaluación y hacer énfasis en su propio contexto. Más adelante se aborda este tema, al igual que los enfoques evaluativos y el proceso para realizar una evaluación.

a través de conocimiento que sea utilizable. Todo en beneficio de la intervención social y del ciudadano.

2.2 Evaluación de programas y políticas públicas: hacia una visión amplia e integral de la evaluación

Limitadamente se ha llamado programa a la clasificación de gastos o asignación de responsabilidades a los gestores, pero un programa es más que eso. Nirenberg, Brawerman y Ruiz (2003: 32) señalan que un programa es “un conjunto coordinado y ordenado de propuestas que persiguen objetivos establecidos en un Plan⁴ y tienden a la atención de problemas específicos relacionados con algunos de los aspectos señalados en el mismo”. Para Subirats (1994: 165) son “las medidas específicas que se toman para llevar a cabo una determinada política”. Así, los planes contienen políticas, y a su vez las políticas dan la pauta para realizar los programas, en este sentido, todos los programas están enmarcados por una política. Concretamente en el ámbito social, Fernández Ballesteros (1992: 477) establece que “un programa es un conjunto especificado de acciones humanas y recursos materiales diseñados e implantados organizadamente en una determinada realidad social, con el propósito de resolver algún problema que atañe a un conjunto de personas”. Los programas públicos pueden tratar temas de educación, cultura, migración, salud, empleo, combate a la pobreza, juventud, vivienda, transporte público, medio ambiente, seguridad, deporte, etcétera.

Para Rutman (1980), los programas son lo que se le entrega al público (beneficios de seguro por desempleo, capacitación, servicios de salud, etcétera) o regulan su comportamiento (regulaciones medioambientales o límite de velocidad) y varían en sus dimensiones (características de la intervención), en su implementación y alcance o ámbito; además, cuentan con servicios de apoyo esenciales para su gestión (financieros, de cómputo, investigación y auditoría). En este sentido, los programas pueden ser evaluados en partes y Rutman (1980) señala tres niveles: el programa, un proyecto del programa, o algunos componentes particulares

4 Por Plan se entiende un documento, generalmente producido por los niveles centrales, que plantea objetivos prioritarios y explicita un conjunto de directivas generales (políticas) en torno a los mismos, las alternativas para alcanzarlos (estrategias) y los medios para obtenerlos (esquema general de asignación de recursos) (Nirenberg, Brawerman y Ruiz, 2003: 32).

del programa. Recomienda que cuando se incluya una variedad de componentes es mejor realizar una evaluación de los componentes específicos más que del programa en su totalidad. Igualmente, un programa debe cumplir con características técnicas y Monnier (1995: 125) señala por ejemplo los objetivos estatutarios generales (lejanos), a medio plazo o inmediatos (operativos); la teoría de acción o estrategia; el campo de acción (delincuencia, vivienda...); los grupos sociales a los que se dirige: destinatarios; recursos humanos, financieros, técnicos o de comunicación; plazos, y evaluaciones.

Para definir la evaluación de políticas públicas y programas, hay consenso en que el emitir un juicio de valor o mérito es la base de toda evaluación, aunque no siempre fue así. Inicialmente, evaluación era solo el “proceso para determinar en qué medida los objetivos de un programa se habían alcanzado” (Tyler, 1949), el énfasis se limitaba en los resultados y la evaluación era una actividad *a posteriori* (Tyler, 1949; Rutman, 1980; Patton, 1987), de ahí que hasta nuestros días perdure esa imagen retrospectiva de la evaluación y esta se asocie a objetivos y resultados. Sin embargo, debemos a Rossi y Freeman (1985: 19) una de las primeras definiciones de evaluación que abarcó el proceso completo de las políticas públicas y programas, y no solo su última fase, ellos definieron la investigación evaluativa como “la aplicación sistemática de los procedimientos de investigación social para valorar la conceptualización y el diseño, implementación y utilidad de los programas de intervención social”.

Stufflebeam (1987: 183) da una definición aún más completa, donde evaluación es “el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los fenómenos implicados”. A partir de aquí, la evaluación empezó a ser entendida, al menos en la teoría, como una actividad que se lleva a cabo incluso antes de que se ponga en marcha el programa, con una visión integral, mostrando preocupación por su utilidad efectiva y práctica, además de vislumbrar el componente plural y participativo que hay en ellas. En años posteriores, las definiciones han coincidido en “una visión integral donde la evaluación se realiza en todas las etapas del diseño de programas, analizando la pertinencia y coherencia de los problemas, objetivos e instrumentos planteados, el proceso de aplicación del programa y sus resultados” (Osuna y Márquez, 2000:3). Se

muestra una preocupación por que la evaluación no se limite exclusivamente a los resultados y a la parte final de los programas o políticas, sino debería abarcar todas las etapas de los mismos y realizar una evaluación para cada etapa.

Las definiciones expuestas son algunas de las más significativas y no marcan una definición exclusiva, ya sea para evaluación de políticas o de programas públicos. Los autores se refieren a ellas de manera indistinta sin notarse una diferencia clara entre la finalidad y metodología que cada una de ellas aplica. En cuanto la manera de llevar a cabo una evaluación se pide apearse a los procedimientos sistemáticos de la investigación social. La finalidad es la valoración integral (diseño, implementación y resultados) para mejorar y ayudar a la toma de decisiones. Sin embargo, una de las diferencias más perceptibles que podemos mencionar es el ámbito al cual se circunscriben, pues mientras un programa tiene acciones concretas y define su ámbito de aplicación y elementos (al menos en teoría); en una política pública es más complicado delimitarlas, ya que al incluir diversas decisiones y acciones gubernamentales, esta tiene una construcción más abstracta, por lo cual resulta más difícil el saber dónde comienza y dónde termina. Cabría decir que en la evaluación, la de políticas públicas es a nivel general y la de programas a nivel particular, pero esta última es indispensable para realizar la primera.

2.3 Funciones de la evaluación. ¿Para qué nos sirve evaluar?

Los tomadores de decisiones tienen acceso a diversas fuentes de información, formales o informales, para basar sus decisiones. Creemos que uno de los mejores medios para producir y proveer información objetiva y relevante para quienes toman decisiones y gestionan programas y políticas públicas, es la evaluación. En este sentido, esta compete con otras fuentes de conocimiento para lograr ser escuchada y tomada en cuenta en todo el proceso decisorio. Es pertinente preguntarnos para qué nos sirve esa información, así como qué utilidad tiene la información recabada. Responder a estas cuestiones es esencial al momento de desarrollar una evaluación, significa también establecer alguna de las funciones básicas que toda evaluación debe cumplir.

La evaluación es una actividad orientada a la acción, y tiene funciones específicas que la distinguen de otras disciplinas o fases del proceso de políticas públicas. Implica tanto la política específica y los programas que

la constituyen, así como las personas que trabajan en las organizaciones encargadas de implementarla, por lo que las funciones a que hacemos referencia abarcan estos dos aspectos. En términos generales, tres han sido las funciones básicas de la evaluación (Stufflebeam y Shinkfield, 1987, 2007; Vedung, 1997, Bustelo, 2001), sin embargo Stufflebeam y Shinkfield (2007) han agregado una cuarta:

1. Mejora. Función generalmente relacionada con evaluaciones formativas. Pretende ser una guía para quienes son responsables de asegurar e incrementar la calidad del programa, en este sentido es prospectiva y proactiva. Al buscar ofrecer calidad en los servicios y un perfeccionamiento en el programa, la evaluación forma parte del proceso diario que afecta a todos los involucrados en el programa, es una retroalimentación que utiliza los resultados para volver a planificar, gestionar y aplicar el programa. En este sentido, la evaluación también se convierte en un instrumento para mejorar el conocimiento y la formación de los implicados en sus propios programas (Osuna y Márquez, 2000).
2. Rendición de cuentas. Relacionado con reportes sumativos. Es un análisis retrospectivo del programa, y aunque en el pasado la rendición de cuentas se refería a poco más que proporcionar datos cuantitativos (número de personas que recibieron el servicio, promedio de tiempo en que los servicios fueron proporcionados a los clientes...), actualmente también se acompaña de información cualitativa que refuerce los datos antes mencionados, abarcando tanto la gestión como los resultados del programa. Como generadora de información (Osuna y Márquez, 2000), la evaluación debe proporcionar la información que solicite no solamente el Congreso y la Administración, sino también la ciudadanía, acerca de las actuaciones del programa en las dependencias públicas que lo gestionan y con la población en general.
3. Iluminar o guiar las decisiones para las acciones futuras. Carol Weiss (1999) refiere lo complicado que resulta que los resultados de las evaluaciones se utilicen. Ella ha llamado “iluminación” al hecho de que ideas originales de la evaluación poco a poco se vayan introduciendo en la conciencia de las personas y cambien su forma de pensar sobre ciertos temas y, por ende, cómo toman decisiones y diseñan políticas. En este sentido, la información obtenida de

las evaluaciones debe servir no solo para guiar los programas, sino también para proporcionar nuevas ideas y maneras de ver y abordar un problema social. Stufflebeam y Shinkfield (2007: 27) consideran que al planear sus estudios, los evaluadores deberían considerar cómo sus hallazgos pueden contribuir a nuevas ideas en problemas de interés a la teoría y a los tomadores de decisiones. Cuidadosamente planeadas, las evaluaciones pueden servir no solo para guiar la operatividad de los programas o la efectividad de los servicios sino también para establecer investigaciones, teorías o preguntas sobre las políticas. Los efectos derivados de esta función no son fáciles de ver a corto plazo.

4. Evaluación como ayuda a la difusión. En esta función, propuesta por Stufflebeam y Shinkfield (2007: 26), uno de los usos de la evaluación es ayudar a los desarrolladores a difundir prácticas probadas y ayudar a los consumidores a tomar decisiones sabias. Aquí el evaluador debe comparar críticamente el servicio o producto con otros competidores. La característica única en la evaluación para la difusión es su enfoque en preguntas de interés práctico para los consumidores.

Aunado a estas funciones básicas, algunos autores mencionan funciones o propósitos mucho más concretos de la evaluación. Fernández Ballesteros (2001) y Osuna y Márquez (2000) señalan una función de control administrativo, de legalidad y financiero. Fernández Ballesteros también hace referencia a que la evaluación justifica decisiones, ya que al hacer una valoración positiva del programa avala las decisiones que se tomaron sobre él. Osuna y Márquez, por su parte, establecen que las funciones de la evaluación son complementarias e interdependientes y, además de las antes mencionadas, establece que la evaluación racionaliza la toma de decisiones públicas, para lo cual hace uso de una perspectiva técnico-económica que no siempre está bien establecida y necesita complementarse con una perspectiva política y social. Otra función es la evaluación como vía de reforma de las prácticas de la Administración, aquí la evaluación es una herramienta en la implementación de las reformas hacia una administración más moderna, con una gestión más eficaz y eficiente que resuelva los problemas de ineficiencia en la gestión de los gobiernos. Subyace la idea de que una sociedad democrática demanda transparencia en la gestión de los recursos públicos.

Una función más se menciona en el Informe para la creación de la agencia estatal de evaluación de la calidad de los servicios y de las políticas públicas en Madrid (2005: 82). Ahí se establece que las evaluaciones no deben entenderse como una prioridad para dirigir críticas al poder, sino como un instrumento para reforzar y legitimar la actuación de los poderes públicos, en ocasiones poco valorada, y servir para tender un puente entre los responsables políticos y los gestores públicos, por una parte, y los ciudadanos, por otra, mediante el cual se fomente la comunicación, transparencia, participación ciudadana e integridad en la gestión de los asuntos públicos.

Stake (2006), por su parte, hace referencia a seis roles habituales de la evaluación profesional de programas, los cuales indican su función y la finalidad central:

1. Evaluar la consecución de los objetivos. Proporciona pruebas cuantitativas del rendimiento de las personas. Además de los objetivos de resultados, también incluye objetivos referidos a los *inputs* o a los procesos. Aparte de concebir los objetivos del programa tal y como están expresados, al evaluador también le corresponde a veces averiguar los objetivos del patrocinador y del personal, así como de otros agentes implicados (por ejemplo, alumnos, resto de profesores, las madres y los padres, y otros miembros de la comunidad), además de los objetivos no declarados. Esta evaluación se aplica principalmente a programas pensados para cambiar a las personas en un sentido u otro, a través de la escolarización o la formación, y lo que pretende es medir sus efectos.
2. Ayudar al desarrollo de la organización. El énfasis está en el funcionamiento interno de la organización, y no tanto en su producción o sus objetivos. La organización funcionaría mejor si la evaluación formal de actividades, procesos, personal y productos, forma parte de la planificación, supervisión, medición y análisis llevado a cabo por la organización. Así, se contribuiría más al desarrollo de la organización, si los propios miembros del personal diseñan, aplican y realizan las valoraciones de la calidad.
3. Evaluar la calidad contextual. Comprende la naturaleza del programa u organización y del contexto en el cual este existe

(histórico, económico, político, estético, sociológico y cultural). Implica reconocer lo bueno y lo malo del evaluando y sus respuestas a ciertas situaciones, para lo cual se requiere de una amplia perspectiva para reconocer las interacciones que se establecen en muchas de sus funciones.

4. Estudiar un ámbito político completo. No solo valora un único evaluando, sino todo un conjunto de ellos, con un afán de generalización. Esto es lo que se conoce también con el nombre de investigación evaluativa.
5. Ayudar a la acción social. Pretende que las conclusiones de la evaluación sean útiles y que el evaluando mejore después de la evaluación, para eso utiliza enfoques participativos.
6. Legitimar el programa. Busca que la evaluación confirme la buena actuación de los directivos o sustente algún tipo de solicitud, sin embargo el evaluador no solo debe enfocarse en las virtudes sino también en los defectos del evaluando.

Con estas funciones más específicas, que bien podrían agruparse dentro de las funciones básicas, hemos visto de manera precisa para qué nos sirve la información recabada de la evaluación. Sin embargo, como contraparte a todas estas funciones, hay ocasiones en que algunas evaluaciones se realizan con fines distintos a los antes mencionados. A estas evaluaciones que persiguen fines “ocultos” o “razones menos legítimas” (Weiss, 1990: 24) las llamamos pseudoevaluaciones (Suchman, 1967; Stufflebeam y Shinkfield, 1987, 2000, 2007; Mejía Lira, 2003), evaluaciones con propósitos encubiertos (Ballart, 1992), con propósitos estratégicos (Vedung, 1997) o evaluaciones perversas (Monnier, 1995). Es probable que recopilen la información de manera rigurosa, pero pueden no revelar las verdaderas conclusiones o si lo hacen solo seleccionan parte de la información. Con base en los autores arriba mencionados, algunos de los fines con los cuales se realizan estas evaluaciones son:

- Ocultar deficiencias y fallos a los superiores.
- Valerse de estas para asumir una postura de racionalidad, anunciándola como una investigación objetiva y científica, ya que una evaluación científica ofrece una apariencia externa favorable y se considera una señal de status profesional.

- Posponer o retrasar una decisión inminente, esperando que con el paso del tiempo el tema de debate o los ataques directos al programa se desvíen hacia otros asuntos y la opinión pública se vaya enfriando hasta la publicación de los informes de la evaluación.
- Adular a un programa débil al justificarlo y elegir deliberadamente para la evaluación solo aspectos que, de antemano, parecen tener éxito y evitar los aspectos aparentemente defectuosos. La evaluación resulta parcial y sesgada.
- Evadir responsabilidades al incorporar a miembros de los partidos de oposición en las evaluaciones, obligándolos a compartir responsabilidades por las acciones impopulares, o en su caso, pretender que la evaluación adopte la decisión en vez del responsable y así evitar toda responsabilidad.
- Conseguir ayuda para obtener, mantener o incrementar un espacio de influencia, poder o recursos económicos. Los problemas que se plantean son los que interesan a quienes solicitan la evaluación y a ciertos grupos que comparten sus intereses.
- Ayudar a quien encarga la evaluación a crear una imagen positiva del programa con datos e información que favorezcan el apoyo de los interesados en el programa y de la ciudadanía en general. O, por el contrario, poner de manifiesto sus resultados negativos con el objeto de utilizarlos como arma política en contra de sus responsables políticos y/o administrativos.

2.4. Propuestas para una mayor utilización de las evaluaciones

La utilidad es el aspecto más importante de las evaluaciones. Una evaluación es eficaz si impacta en las decisiones tomadas. Partimos en principio de que la evaluación produce y provee información objetiva y relevante, susceptible de guiar e influir en las políticas gubernamentales. Sin embargo, ha sido evidente la influencia limitada de los resultados o las recomendaciones en los procesos de toma de decisiones. El uso de la evaluación fue durante mucho tiempo un tema al cual no se le prestó importancia, se pensaba que la utilización de los resultados y las recomendaciones eran algo que sucedía en automático. En la década de los setenta y principios de los ochenta surgió la preocupación por el uso de los resultados de la evaluación, a raíz de que a pesar del elevado costo que

conllevaban, los responsables políticos o gestores daban poca atención a sus resultados debido a:

- El modelo racional en el cual se fundamentan muchas evaluaciones.
- La escasez de información básica relativa a las actuaciones o personas a evaluar.
- La falta de tiempo para desarrollar todas las fases y actividades del programa.
- La entrega tardía de los resultados.
- Los resultados que no respondían a las necesidades de los decisores, trayendo como consecuencia su irrelevancia y nula utilización.
- La forma de presentación de los informes y su lenguaje excesivamente técnicos.
- La elaboración de los informes (no se hacían pensando en los decisores o gestores).
- El cambio (durante el proceso de evaluación) del decisor que pidió el estudio y el que recibía los resultados.
- Las resistencias al cambio en la organización.

Los evaluadores debían tratar con el poco uso de los resultados. Patton (1986) fue de los primeros en interesarse en eso, y más que preocuparse por determinar la calidad del programa, su preocupación fue que la evaluación ayudara a mejorar el funcionamiento del programa (modelo de evaluación dirigido a la utilización ⁵). Monnier (1995: 142) refiere que la utilidad social de una evaluación depende de las condiciones de apropiación de sus conclusiones por parte de los actores sociales a los cuales va destinada, así como de la legitimidad que tiene ante quienes aplicarán las mejoras sugeridas. Si esto es así, la comunicación constante entre el evaluador y las personas interesadas en la evaluación es una necesidad, es decir, la evaluación y el evaluador deben responder a las necesidades de información de los decisores, gestores o quienes solicitan este trabajo. Si lo que se busca es impactar en la gente real, en las organizaciones y en la sociedad, el evaluador debe ver que sus posibles audiencias estén abiertas a usar y ser influidas por los resultados evaluativos.

5 Este modelo –Evaluación centrada en la utilización– se expone más adelante, dentro de los enfoques evaluativos más relevantes para inicios del presente siglo.

El informe y los resultados de la evaluación no son interés exclusivo de un solo actor, sino existen distintas partes interesadas en esos resultados, por eso es evidente que distintas audiencias requerirán diferente información. Carol Weiss (1990) menciona la variedad de usuarios que pueden tener las evaluaciones, por ejemplo:

- Los encargados de tomar decisiones requerirán información acerca de la continuación o terminación del programa, si deben seguir con la misma estrategia o no.
- Los directores del programa estarán interesados en saber cuál estrategia tiene mayor o menor éxito, cuál alcanza los resultados de la manera más eficiente y económica.
- Los miembros del personal del servicio que tratan con los destinatarios, querrán saber cuestiones más prácticas y del trabajo cotidiano.
- Quienes financian la evaluación también requieren información para saber si continúan o no con el apoyo económico al programa.

Esta diversidad de actores y usuarios potenciales de la evaluación pone en evidencia el reconocimiento de “factores políticos y el planificar las evaluaciones teniendo en cuenta a sus posibles usuarios e incluso su participación activa en el proceso evaluador” (Ballart, 1992: 153). Resalta así la diferencia que señala María Bustelo entre el solo hecho de proveer información y la participación activa de los distintos agentes en los procesos de evaluación. Bustelo (2001: 243) destaca que preguntar la opinión o qué recomendarían hacer los agentes a través de encuestas, entrevistas o grupos de discusión, no significa participar en el proceso de evaluación, sino ofrecer información o convertirse en información que será utilizada en el proceso sistemático de evaluación. La participación real y activa de los distintos agentes en la fase de valoración y establecimiento de las recomendaciones de una evaluación, implica que se les ofrezca toda la información sistematizada y analizada, y se les pidiera, no su opinión, sino una valoración de dicha información y sugerencias a raíz de la misma.

Si la rendición de cuentas es una de las funciones de la evaluación, aunado a que los programas públicos y la mayoría de las evaluaciones se financian con fondos públicos, es un hecho que los resultados de las evaluaciones no solo deben llegar a los miembros del Congreso y organismos

públicos, sino también a los destinatarios del programa, a los académicos y profesiones, y al público en general.

Vedung (1997) hace tres propuestas para fomentar la utilización ⁶ de los resultados de las evaluaciones:

1. Perspectiva centrada en la difusión (que los receptores conozcan los resultados). La poca utilización se debe a barreras en la comunicación entre quienes producen y quienes consumen la evaluación, así como a la elaboración y difusión de los informes. La transmisión de resultados se puede mejorar a través de documentos, folletos y resúmenes orales, que se pretende lleguen, de una manera amena, al mayor número posible de audiencias. Los informes deben ser concretos y breves, deberían concentrarse en un solo asunto importante, y dentro de los mismos informes debería haber un resumen breve y concreto, utilizar gráficos en el apartado de resultados, resaltar los resultados cruciales, reducir al mínimo las cuestiones metodológicas y ponerlas en apéndices, hacer llegar los principales resultados y recomendaciones a las audiencias de forma continua, y antes de que se haya redactado el informe final e incluir recomendaciones para la acción. Se recomienda comunicar los resultados en persona: hablar de manera breve y frecuente, contando historias y anécdotas para ilustrar las explicaciones. Es importante que exista un método de conexión mediante el cual la difusión se institucionalice en grupos formales, comisiones o algún sistema de documentación permanente.
2. Estrategia centrada en la producción (proceso de la evaluación). Las evaluaciones se desacreditan o se archivan porque son irrelevantes o inadecuadas, en este sentido, el proceso de evaluación se debe adaptar para responder a las demandas y deseos de los receptores. Por ende, se propone y apoya la participación activa de los usuarios potenciales de la evaluación, lo cual permite un aprendizaje durante el proceso y un mayor compromiso con los resultados. Las

⁶ Respecto del uso de los resultados de las evaluaciones en México, Cardozo (2013 y 2015) realiza un análisis exhaustivo acerca del uso de los resultados de las evaluaciones solicitadas por el Coneval y Evalúa CDMX. Otro trabajo relevante es el desarrollado por Jaqueline Meza (2017) en el marco de su tesis doctoral titulada “Principales factores asociados al uso instrumental de la evaluación externa de los programas presupuestarios federales en México, periodo 2007-2013”.

evaluaciones deben acomodarse a las distintas fases de madurez del programa: diferentes fases requieren diferentes valoraciones.

3. Metaevaluación. Se refiere a tres operaciones diferentes: valoración (evaluación) de otra valoración que esté en curso o que haya finalizado; procedimiento para resumir y sintetizar los resultados de varias valoraciones del mismo programa o de un conjunto de programas similares, y la evaluación de la función valorativa general de una organización o de una unidad.

Monnier (1987 en Meny y Thoenig, 1992: 213) hace referencia a las condiciones preliminares que se debe satisfacer para que la difusión de la evaluación no se haga difícil:

- Relativa independencia de los organismos de evaluación frente a los prescriptores. Los organismos pueden ser en parte del gobierno o totalmente autónomos (asociaciones, centros de investigación, etcétera).
- Capacidad pluridisciplinar de los evaluadores.
- Transparencia suficiente de la información administrativa, que permita el acceso a las fuentes.
- Pluralidad de órganos de evaluación que permita la emulación entre expertos y su utilización por los decisores como bazas en la negociación.

Así como los resultados de la evaluación pueden llegar a distintas audiencias y tener distintos usuarios, la evaluación también puede tener distintos usos. El uso, como tal, de la evaluación ha sido tratado por diversos autores: Weiss, 1977; Rich, 1977; Levinton y Hughes, 1979, 1989; Rossi y Freeman, 1989; Subirats, 1994; Vedung, 1997. Entre los usos más comunes destacan:

- Uso conceptual. Los resultados de la evaluación se incorporan gradualmente en el marco de referencia global del usuario, influyendo sobre el pensamiento (Weiss, 1979) pero no en la acción. Es decir, el receptor le ha prestado atención a la evaluación, se ha interesado por esta, la ha comprendido y, por su mediación, ha adquirido algunas habilidades que no tenía antes, pero no han

puesto en marcha acción alguna o actuado en base a esos resultados (Vedung, 1997). Busca más la capacidad general de influenciar la forma de pensar en los problemas y temas planteados, sin que de los resultados surja una acción inmediata o cambio (Subirats, 1994).

- Uso instrumental. Son las acciones efectuadas por los tomadores de decisiones. Es cuando de los resultados obtenidos de la evaluación se derivan acciones inmediatas o cambios en los programas de actuación (Subirats, 1994).
- Uso interactivo. Son muchos los actores que intervienen en el programa evaluado, y su implicación activa en la interpretación de los datos favorece la expresión de nuevos planteamientos, además, lleva a las partes a reconsiderar sus posturas con respecto del problema social del programa evaluado. Este uso implica la aplicación de conocimientos basados en la evaluación, junto con nuevos datos basados en la investigación y otras formas de antecedentes, tales como el sentido común, la sabiduría convencional, la intuición y las propias experiencias de primera mano de los destinatarios (Weiss, 1979: 206s).
- Uso persuasivo. Los resultados de las evaluaciones dan argumentos que se usan para la defensa o el ataque de un cierto programa, organismo de actuación o posiciones políticas.
- Uso como legitimación. La evaluación se usa para justificar posturas previamente establecidas y fundamentadas en otras razones, ya sean políticas, electorales, etcétera. También se utiliza para ignorar o desacreditar aquellos resultados que se consideran perjudiciales.
- Uso táctico. La evaluación se usa para ganar tiempo o eludir responsabilidades. El hecho más importante no lo constituyen los hallazgos, sino el que esté en marcha una evaluación, por lo tanto, lo que se utiliza es el proceso de evaluación y no el resultado mismo (Vedung, 1997).

En resumen, distintas audiencias tendrán distintas necesidades de información, así que como lo señalan Stufflebeam y Shinkfield (2007), es un requisito básico que el evaluador sea sensible a las necesidades de los clientes respecto de la mejora del programa, y que tenga habilidades para comunicar los resultados de la evaluación a través de presentaciones escritas

y orales, claras y coherentes. Agregan que para un uso de la evaluación se debe involucrar a todas las partes interesadas (*stakeholders*⁷) en la misma, identificar a las audiencias y sus usos potenciales sobre los hallazgos de la evaluación. Ante esto, no queda más que una constante comunicación.

7 El uso de la palabra *stakeholders* es muy común dentro de la literatura de evaluación de políticas públicas y programas. Sin embargo, en este libro parece más apropiado el uso de su significado y equivalente en español: partes interesadas. Refiere a todas las personas afectadas por la evaluación, es decir todos los actores con intereses en la evaluación o programas. Su implicación puede ser directa o indirecta.

3. El innegable componente político en la evaluación. Actores y contexto

La democracia en que vivimos, y su respectiva calidad, implica la participación de nuevos actores en los procesos de decisión, implementación, y resultados de la acción de los gobiernos. En cada escenario hay necesidad de información, y la evaluación puede servir a esos distintos actores, quienes de acuerdo con sus intereses, ayudarán u obstaculizarán el desarrollo de la política o programa público. Si fuera cierto lo anterior, se asume que la evaluación se lleva a cabo en un contexto político y por lo tanto la evaluación está permeada por un componente político. Incluso, desde la ciencia política la evaluación se concibe como una actividad intrínsecamente política, en el momento que en ella se ven involucrados actores con intereses contrapuestos, políticamente relevantes, que luchan por obtener una mayor cuota de poder e influenciar en el proceso evaluativo, con el propósito de modularlo de acuerdo con sus preferencias (García Sánchez, 2009). Pero este componente político y los actores involucrados se han considerado como perjudiciales para la evaluación –incluso hoy en día–, por lo que no era recomendable implicarlos en la tarea evaluativa. Al componente político se le percibe como negativo, da la idea de problema, manipulación o incluso mentira, sin embargo debemos reconocer que eso es parte de la naturaleza en la interacción humana.

El hecho básico de identificar y establecer valores y estándares, implica ya un componente político, pues entre las partes involucradas tendrán que definir qué valores imperarán sobre otros. El hecho de establecer el uso de los hallazgos de la evaluación y que esta influya en las políticas y programas,

es también un hecho político, así como involucrar a los distintos agentes durante el proceso evaluativo y establecer sus necesidades evaluativas. Fueron, en cierta medida, los resultados desalentadores en las evaluaciones los que demostraron que los actores, además de hacer el proceso más complejo, podían ayudar a obtener mejores resultados y hacer más útil la evaluación. A estos actores con intereses en el programa o la evaluación se les conoce como *stakeholders*, es decir, son las partes interesadas.

Para Stake (2006) los actores implicados son a quienes afecta de algún modo el programa, son un colectivo de actores heterogéneo que incluye a las partes beneficiarias y a las perjudicadas. Subirats (1995) señala que el hecho de que todo programa o acción que vaya a evaluarse tenga sus propios políticos responsables de tomar las decisiones, patrocinadores, personal administrativo que participa en él, clientela y grupos de interés que en cierta medida han sido afectados por él, ilustra la estrecha relación entre el escenario político y el proceso de evaluación. Por su parte, el público para un trabajo de evaluación está integrado por las personas que prestarán atención a los resultados y hallazgos de los evaluadores: los gestores y patrocinadores del programa, otros evaluadores y algunos actores implicados en el programa.

De acuerdo con Jody Fitzpatrick (2011), es esencial que exista una comunicación entre el evaluador y las partes interesadas, para que todos conozcan los propósitos y los beneficios de la evaluación, así como para negociar los aspectos relevantes de la misma. Al final, el evaluador y las partes interesadas se complementan, pues los primeros son los expertos en los métodos (aspecto técnico), pero los segundos tienen la experiencia en el programa que se está evaluando. No todas las partes interesadas querrán participar con la misma intensidad, pero corresponde al evaluador identificarlos y conocer sus intereses y opiniones, sin tener que favorecer los intereses de algún grupo en particular.

Para Cronbach (1980), son cinco los grupos que influyen en el proceso de dar forma a una política: los responsables a nivel de formulación de la política, los responsables a nivel de programa, 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.

La evaluación se vuelve entonces, no solo una tarea técnica y metodológica, sino también una tarea de relaciones interpersonales que

se dan en un contexto político con conflictos y consensos. “Durante el proceso evaluativo habrá rivalidad de opiniones, diferentes necesidades, necesidades que compiten e intereses opuestos, lo que garantiza un desacuerdo acerca de las reglas bajo las cuales la gente convive. Por el otro lado, la gente reconoce que para influenciar estas reglas o asegurar que las normas se respeten, ellos deben trabajar con otros” (Vestman y Conner, 2006). Si no se considera el factor político, existe el riesgo de que, como lo menciona Fitzpatrick (2011), muchas evaluaciones impecables en aspectos técnicos fracasen por insensibilidad interpersonal y poca consideración de los aspectos políticos, pues los evaluadores necesitan entender que las partes interesadas tienen diferentes perspectivas, intereses y preocupaciones acerca del programa y de la evaluación, así como entender también el aspecto político en el cual se desenvuelve la evaluación. Con esto, además de habilidades técnicas, el evaluador necesita habilidades en relaciones humanas para trabajar con diferentes grupos dentro de un contexto político.

Weiss también ha discutido este tema (1973, 1991) y señala al menos tres formas en las cuales la evaluación y las políticas están unidas:

1. Las políticas y programas con las cuales los evaluadores tratan son ellas mismas producto de una decisión política.
2. La evaluación es una herramienta de insumo en la toma de decisiones, su reporte entra en la arena política, donde la evaluación provee información.
3. La evaluación en sí misma tiene una postura política. La evaluación, por su naturaleza, hace declaraciones implícitamente políticas, por ejemplo los retos que legitiman las metas de ciertos programas o estrategias de implementación.

Weiss (2004: 121) añade que si queremos aprender a desarrollar programas con mayor sentido de la realidad, reducir las discrepancias entre los deseos y la acción y tratar con mayor eficacia los problemas sociales.

El evaluador debe comprender qué es lo que ocurre en las complejidades políticas y sociales de los programas, puesto que todo programa se realiza en un marco o ambiente social que tiene consecuencias para su eficacia. El contexto primario es el de la organización que patrocina y lleva a cabo el programa; los marcos sociales más amplios son la comunidad, que también afecta a los programas y sus consecuencias. Así como el programa está

inserto en un contexto social, la persona que participa también lo está, y llega con vínculos, raíces, creencias, valores, hábitos, conductas e ideas (Weiss, 2004). Cronbach (1980) ve la evaluación como una actividad esencialmente política, mediante su influencia en las decisiones políticas y la formulación de la política, la idea principal es entender el problema y el programa, el contexto político y sobre todo qué procesos contribuyen a los resultados esperados.

Siguiendo con Fitzpatrick (2011), la evaluación es un acto político al momento que los evaluadores trabajan en un contexto político. Si la evaluación está relacionada con guiar o influenciar las políticas del gobierno, es un hecho que los evaluadores trabajan con actores cuyas preocupaciones son distintas, también están interesados en guiar e influenciar las políticas y compiten entre ellos por recursos, poder y liderazgos, y en este sentido la evaluación es una poderosa herramienta para pelear por recursos para su grupo o programa. De igual manera, las evaluaciones tienen lugar en organizaciones donde las relaciones complejas entre muchos grupos existen, donde la más básica evaluación puede molestar o cambiar esas relaciones, y puesto que la evaluación involucra un juicio acerca de la calidad de lo que se evalúa, la evaluación es altamente política. Si a esto agregamos que a menudo las partes interesadas presionan a los evaluadores para que la evaluación luzca bien o mal, o su programa parezca exitoso, el componente político es más que evidente, y no solo durante el desarrollo de la evaluación, sino desde su diseño y hasta la presentación de sus resultados.

Vestman y Connors (2006 y en Fitzpatrick, 2011) describen tres posiciones en las cuales los evaluadores pueden interactuar con el sistema político:

1. El evaluador como valor-neutral. Trata de proteger o separar la evaluación de la política para mantener su legitimidad y objetividad. El evaluador es metodológico racional, recolecta datos y los da a las partes interesadas y trabaja para mantener separada e independiente la evaluación y conservar así su “objetividad”.
2. El evaluador como valor-sensitivo. Trabaja para mantener los aspectos técnicos de la evaluación y la provisión de la información como separados de la política, pero reconoce que otros elementos de la evaluación, en particular los juicios, tienen en cuenta aspectos éticos y alientan los valores democráticos, requieren que el evaluador aprenda y se llegue a involucrar en el ambiente político.

3. El evaluador como valor crítico. Cree que los valores son inextricablemente parte de la política y es importante que el evaluador llegue a estar implicado en la política para articular activamente esos valores. Aquí, el evaluador “ve la política como algo integrado en nuestra vida diaria” y, por lo tanto, “no puede haber separación entre evaluación y política y por lo tanto ningún valor neutral o posición operativa tomada por el evaluador” (Vestman y Connors, 2006: 235). El evaluador toma un rol activo al considerar qué es bueno para el público.

Con base en lo anterior, la técnica no es suficiente para realizar evaluaciones útiles en contextos reales. Si se quiere que la evaluación sea usada por quienes la han solicitado y cumpla con sus funciones, es un hecho innegable que se debe involucrar a todo aquel que tenga algo que ver con el programa y la evaluación. Tener en cuenta el contexto político, no solo a nivel micro sino macro, es importante, pues las decisiones generales o federales afectan las decisiones sobre evaluación de los involucrados en los programas a nivel local. Y lo más importante: el evaluador y clientes deben estar abiertos a no ver a la evaluación como un simple instrumento técnico para obtener información y datos para la toma de decisiones, sino considerar que la política permea la evaluación y los valores de los evaluadores y clientes, sus antecedentes o su género también llegan a ser parte de la operación explícita o implícita de la política en evaluación (Vestman y Connors, 2006).

4. La evaluación a través de los años. De la evaluación como medición de resultados a la evaluación como aprendizaje organizacional

La evaluación de políticas públicas y programas tuvo su origen y desarrollo en los Estados Unidos, debido principalmente al carácter de obligatoria que la legislación federal estableció para los programas que ella misma impulsaba, financiando al mismo tiempo dichas evaluaciones. Se autorizó a comités del Congreso llevar a cabo los estudios, contratarlos o solicitar a otras agencias del gobierno el llevarlos a cabo.

Uno de los primeros antecedentes en evaluación se remonta a inicios del siglo XX, en el año de 1904 con el trabajo de Alfred Binet, quien se considera un pionero en la evaluación. Binet, a solicitud del Ministerio de la Instrucción Pública Norteamericana, desarrolló un test psicológico conocido como Test de Inteligencia Binet. Dicho test pretendía detectar a los niños con capacidades intelectuales bajas, a fin de separarlos de sus compañeros con inteligencia normal, y dotarlos de una educación especial. El objetivo del Ministerio era mejorar la instrucción pública mediante una selección de alumnos. Este test fue una de las primeras herramientas experimentalistas al alcance de los poderes públicos. El trabajo realizado por Starch y Elliot en 1912 se acercó más a lo que conocemos como el enfoque experimentalista en evaluación. Se desarrolló en el campo de la formación de profesores de matemáticas y consistía en una comparación entre dos grupos de profesores, uno de los grupos había seguido un entrenamiento especial y el otro era el grupo de control.

Sin embargo, fue en los años treinta cuando se analizaron los primeros programas educativos y de salud pública. Uno de los trabajos que marcó los inicios en la evaluación y que influyó de manera notable la conceptualización de la evaluación educativa de los años siguientes fue el realizado por Ralph Tyler en 1932. Elaboró un programa piloto con alumnos de bachillerato, siguiendo su formación hasta los estudios universitarios, comprobando si habían tenido el mismo éxito en sus estudios universitarios que sus compañeros procedentes de un sistema escolar tradicional. Los objetivos del programa eran su punto de referencia y debían expresarse en términos conductuales u observables ⁸, pues servirían de base para la elección de las variables dependientes y de los criterios utilizados con los cuales se contrastaban los resultados. Este era un enfoque muy riguroso y evaluación y medida eran sinónimos ⁹.

Los investigadores llamaron a su trabajo “investigación evaluativa”, por la aplicación de los métodos de la investigación social. Las décadas cuarenta y cincuenta fueron de relativa inactividad, fue a finales del sesenta con los programas de la Gran Sociedad que la evaluación se presentó con mayor fuerza. En los años sesenta, después de la Segunda Guerra Mundial y bajo el mandato de los presidentes Kennedy, Johnson y Nixon, se iniciaron grandes programas innovadores en educación, salud, vivienda, etcétera. destinándose fuertes cantidades de dinero a lo que se llamó Guerra contra la pobreza. El gobierno no contaba con suficientes empleados capacitados para hacer frente a estas nuevas innovaciones, no había retroalimentación sobre el rendimiento de los programas, y técnicas que habían funcionado en programas pequeños resultaban inadecuadas para programas de gran escala. Los problemas públicos parecían ser insolubles y los objetivos iniciales se perdían a medida que el programa se implementaba, dando paso, incluso, a nuevos problemas.

Se empezó a dar cuenta de que era un error el énfasis en la definición de problemas y toma de decisiones, dejando de lado y considerando poco problemático la ejecución y la consecución de objetivos; una cosa era un programa trazado de manera perfecta y otra su aplicación práctica. Los recursos se limitaban y la eficacia de los programas se cuestionaba cada vez más. Aquí, la evaluación entró como una forma de analizar los programas

8 Se destaca que esta orientación era consecuencia de la hegemonía del conductismo en la ciencia política de esa época.

9 Cabe recordar que para Tyler, evaluar era el “proceso para determinar en qué medida los objetivos de un programa se habían alcanzado” (Tyler, 1949).

de manera sistemática y retrospectiva, la atención ya no se centraba exclusivamente en la fase de decisión; el interés ahora era entender todo el proceso de las políticas públicas. En este contexto, la evaluación era cada vez más necesaria, dando paso en los gobiernos a un periodo de rendición de cuentas y ya no tanto de expansión.

El crecimiento de la evaluación también debe ser visto dentro de un contexto de reformas en la Administración Pública (Rutman, 1980). Con la intención de hacer la planeación, el presupuesto y los procesos de gestión más racionales, surgió una variedad de enfoques sistemáticos, como la Presupuestación por Programas, el Presupuesto Base Cero, la Dirección por Objetivos, el Análisis Costo-Beneficio y la experimentación con prospectiva de programas y políticas. Así por ejemplo, la Presupuestación por Programas u Organización por Objetivos, consistía en la fijación de objetivos, selección de alternativas posibles, elección de indicadores de gestión y creación de mecanismos de cuantificación como medida de éxito o de consecución del objetivo (Subirats, 1992). Aquí la evaluación de programas fue un vínculo para proveer la metodología y/o la información necesitada para implementar estas reformas ¹⁰.

Los programas de la Gran Sociedad ¹¹ generaron entre sus directores y tomadores de decisiones la inquietud sobre cómo mejorar los programas y conocer estrategias para alcanzar las ambiciosas metas. El Congreso, por su parte, quería información para continuar con la financiación. Con esto se incrementó la demanda de evaluaciones y el gobierno federal expandió el gasto destinado a estas con el propósito de supervisar los nuevos programas sociales. Ante estas condiciones se requería de nuevos enfoques, métodos, estrategias, y personal calificado, pues era evidente la falta de trabajo teórico y metodológico sobre la evaluación. Aparecieron entonces publicaciones, revistas especializadas, libros y artículos sobre teorías, modelos evaluativos

10 En México, actualmente, con el impulso de las reformas neoliberales sustentadas en los principios de la Nueva Gestión Pública, se ha dado un impulso a la evaluación enfocada en el desempeño del gobierno y en la gestión y el presupuesto para resultados. La evaluación se basa en el Sistema de Indicadores de Desempeño que ofrece la información para los procesos de planeación, programación y presupuestación del gasto público federal.

11 La Gran Sociedad fue una reforma social emprendida en los Estados Unidos por el presidente Lyndon B. Johnson, para combatir la pobreza. La reforma cubría áreas como educación, salud, vivienda, migración o discriminación. A pesar de lo ambicioso, los resultados no fueron los deseados, lo que cuestionó la eficacia de los programas.

y metodologías por parte de diversos evaluadores ¹². En los años setenta la evaluación se convirtió en una industria en crecimiento (Rossi y Freeman, 1993) y surgió la época del profesionalismo (Stufflebeam y Shinkfield, 1987) debido a la exigencia de especialistas que condujeran evaluaciones útiles. Apoyados por el gobierno, las universidades ofrecieron cursos sobre evaluación y demás temas afines. Se resaltó también la importancia de realizar juicios acerca del mérito y valor. La evaluación ya no era vista como una simple medición de si los objetivos eran alcanzados.

En la década de los ochenta se crearon las primeras asociaciones profesionales que contribuyeron a la formalización de la evaluación como una profesión con códigos éticos para guiar la práctica y estándares para juzgar sus resultados, así como para el entrenamiento e intercambio de ideas y aprendizaje. Se creó el *Joint Committee on Standards for Educational Evaluation* (1981, 1988), el cual elaboró el texto *Standards for Evaluations of Educational Program, Projects and Material* que establecía una serie de normas (treinta) para la evaluación de programas, proyectos y materiales educativos; aconsejaba que los clientes y evaluadores cooperaran entre sí para conseguir que la evaluación fuera útil, factible, ética y exacta. En esta década, las evaluaciones ya no eran solo federales, también se realizaban a nivel local, con lo que nuevos enfoques surgieron (evaluaciones formativas, evaluaciones centradas en la utilización, evaluaciones que consideraban a las múltiples partes interesadas y con métodos más cualitativos).

Los años noventa vieron cómo las evaluaciones se llevaban a cabo en diversos países, diversos contextos y usando diversos enfoques y métodos.

12 Muchos de estos trabajos son referentes básicos e influyentes, aún hoy en día. Baste con mencionar como ejemplo a Suchman (1967), quien publicó un texto revisando los diferentes métodos de evaluación; Campbell (1969) argumentó más experimentación social al examinar la efectividad de los programas; el libro de Campbell y Stanley (1966) en diseños experimentales y cuasi experimentales fue muy influyente; Scriven (1967), Stake (1967) comenzó a escribir artículos acerca de las teorías y prácticas evaluativas; Wholey y sus colegas (Wholey *et al.*, 1970) reconocieron los aspectos políticos que conducían a la evaluación dentro de las organizaciones; Stufflebeam (1972) respondiendo a la necesidad de evaluaciones que fueran más informativas a los tomadores de decisiones desarrolló el modelo CIPP; Stake (1975a) propuso la evaluación responsiva, moviendo a los evaluadores de los paradigmas dominantes de la experimentación y ciencia social; Guba y Lincoln (1981) propusieron la evaluación naturalista, dirigiendo del debate sobre los méritos de los métodos cualitativos y cuantitativos (Fitzpatrick, 2011).

Se puso especial atención en la institucionalización de la evaluación (Sufflebeam, 1997 en Sanders, 2007: 254), lo que supuso un compromiso de las organizaciones hacia la evaluación. Muchas organizaciones en diversos sectores hicieron este compromiso a través de la acreditación, auditorías, informes de las políticas, guías, asignación de personal, o alguna otra evidencia tangible de inversión en la evaluación (Sanders, 2002: 254). Esta época puso énfasis en los resultados y la rendición de cuentas, los programas o instituciones buscaban el logro de estándares establecidos y continuar vigentes, la evaluación era tanto sumativa como formativa. Se busca en la evaluación un medio para garantizar calidad, competitividad y equidad en la entrega de servicios. Esta visión aún continúa en nuestros días y se ha extendido a países de América Latina.

En la actualidad, la evaluación se caracteriza por la preocupación de la calidad de las evaluaciones. Han surgido más agencias que establecen estándares para la evaluación de programas como los Principios Rectores para los Evaluadores, desarrollados por la Asociación Americana de Evaluación (2003) o los Estándares de Auditoría Gubernamental (2003) de la Oficina de Rendición de Cuentas Gubernamental de los Estados Unidos. Esta construcción de calidad en la evaluación incluye una serie de prácticas dirigidas hacia su mejora y su uso en la organización, y llevarla como la corriente principal, esto último significa que la evaluación está en el primer plano del pensamiento y comportamiento de la organización (Sanders, 2002). La evaluación se ha extendido a todas las áreas y niveles del gobierno, lo cual ha generado que mucha gente (directores, supervisores, empleados de organizaciones y escuelas...) tenga responsabilidades en evaluación como parte de su trabajo. Sanders (2002) menciona que la posición de la evaluación ha mejorado, pero esta todavía no tiene una naturaleza arraigada en las organizaciones y su influencia como proceso aún es limitada. Habla sobre *mainstreaming* (corriente dominante) al referirse al “proceso de hacer de la evaluación una parte integral en las operaciones diarias de las organizaciones” (2002: 254), en vez de ser puesto al margen del trabajo la evaluación se convierte parte de la rutina en el trabajo ético de las organizaciones si es una corriente dominante.

En este sentido no podemos dejar de mencionar el rol de la evaluación en el aprendizaje organizacional, que actualmente está en discusión y surge a raíz de la preocupación de los evaluadores por incorporar a las distintas partes interesadas de la evaluación y de los programas en las organizaciones

para usar la información evaluativa. Este término se refiere no a los cambios ocurridos por la información de la evaluación, sino a lo que las partes interesadas aprenden por participar en el proceso de evaluación, con lo cual el rol de la evaluación se amplía. Este es un tema que aún está en desarrollo y como Sanders menciona (2002), lo que busca es la eficacia de la organización (que se traduce, por ejemplo, en mejores servicios) a través de muchas mentes trabajando juntas, compartiendo información y hacia una meta en común, lo cual se logra cuando hay una cultura del valor de la evaluación y cuando sus labores hacen continuo uso de la evaluación. Añade que no es suficiente poner más recursos en la evaluación si esta no impregna la cultura de la organización.

Estos hechos históricos acerca de la evaluación señalan fases que han ido marcando una evolución en la forma de ver y hacer la evaluación, desde una evaluación centrada en los objetivos hasta una que apela al pluralismo dentro de la misma. Los autores coinciden en dividir esta evolución en cuatro etapas (Guba y Lincoln, 1989; Shadish, Cook y Levinton, 1991; Menny y Thoenig, 1992; Tamayo Sáez, 2002):

1. La evaluación como medición de resultados.

Tuvo su inicio dentro de la educación, psicología y sociología. Considera a la evaluación y a la medición como sinónimos, hace uso de las técnicas y herramientas más objetivas y científicas posibles. La unidad de análisis eran los individuos y el análisis se limitaba a los resultados, los cuales a menudo producían efectos inesperados que el propio análisis no explicaba.

2. El juicio sobre los objetivos.

Se amplía la unidad de análisis hacia los programas públicos, aunque continua limitándose a los objetivos del mismo. Trata de obtener conocimiento para rediseñar los programas y que su funcionamiento sea compatible con los objetivos declarados (Tamayo Sáez, 2002). La evaluación se convierte en un fin en sí misma.

3. La innovación social por la experimentación.

El énfasis se pone en la construcción y la demostración de las relaciones causa-efecto entre programas (variables independientes que se pueden manipular) y los cambios en el programa (variable dependiente en la que se pueden observar los efectos de la manipulación). El método utilizado es la experimentación y las técnicas cuantitativas. Los evaluadores van más allá de un juicio y son quienes generan las soluciones, convirtiéndose en actores fundamentales del proceso.

Tamayo Sáez (2002) distingue tres aproximaciones teóricas en esta fase:

- a. El hincapié en las cuestiones metodológicas y en el rigor científico de Scriven y Campbell, quienes se preocupan por mostrar causalidad y por abstraerse de la influencia de agentes sociales que intentan utilizar la evaluación como instrumento para sus intereses. Entienden la experimentación como la comparación de distintos programas para llegar a los resultados. Los juicios de valor se basan en hechos y no en posiciones subjetivas, dan por sentado el uso de los resultados de la evaluación, asumiendo que los decisores adoptaran los resultados y los utilizaran como base en sus decisiones. No hay conexión entre evaluadores y decisores.
- b. El uso de la evaluación y sus resultados planteado por Scriven y otros autores. El evaluador, además de ser un experto metodológico, debe ser un experto comunicador dando la información adecuada, de acuerdo con las necesidades de los distintos clientes y usuarios de la evaluación. Las relaciones entre los expertos y grupos de interés ya no se ven como perniciosas, sino todo lo contrario, y empieza a reconocerse que los programas se ven influidos por cuestiones políticas. Surge el debate entre cuantitativistas y cualitativistas.
- c. Superación de las contradicciones anteriores al introducir aspectos políticos en sus diseños y abogar por una cierta diversidad metodológica en la evaluación. Pero a pesar de esto el paradigma científico prevalece, por lo cual no resulta sencillo introducir el pluralismo de valores. Rossi y Cronbach son los principales representantes.

1. La función de la evaluación como actividad política.

A partir de los años ochenta surgió un enfoque alternativo que vio a la evaluación como negociación (Stake, 1975; Guba y Lincoln; 1989) y propuso que los actores sociales participaran o evaluaran los actos de gobierno. El elemento central de la evaluación es que todos los actores con influencia en el proceso de las políticas realicen aportaciones en todas las etapas del proceso de la evaluación. Así, la evaluación es resultado de un proceso de debate social y no solo de un proceso estrictamente técnico, debiéndose dar respuesta a los intereses de los ciudadanos y de los actores presentes en las políticas o programas. La evaluación es un proceso de aprendizaje-enseñanza, donde lo que más importa es generar conocimiento aplicable y no tanto el hecho de que una política haya generado, o no, los resultados esperados. Aquí el evaluador se convierte en mediador.

5. Los métodos cuantitativos y cualitativos. La importancia de los paradigmas en la evaluación

En este texto apelamos a la realización de evaluaciones integrales que abarquen todo el proceso del programa, y si bien la evaluación se diferencia de la investigación ya se ha mencionado que para realizar una evaluación se utilizan los métodos de las ciencias sociales o de la investigación social, y los juicios de valor y recomendaciones deberán basarse en una investigación sistemática y rigurosa. Dos han sido las tendencias que han dominado la realización de las evaluaciones: la corriente cuantitativa y la corriente cualitativa, cada una por su parte ha defendido ser el mejor método en la investigación social o evaluación. Sin embargo, quienes han centrado el debate en las ventajas e inconvenientes técnicos de ambos enfoques concluyen que es posible su integración o combinación, la cual se ha reflejado en los enfoques evaluativos que muestran su validez en este siglo. No obstante esta dicotomía encierra problemas más allá del técnico, existen implicaciones en los planos ontológico, epistemológico y metodológico, es decir, a nivel de paradigma.

Antes de avanzar conviene señalar que un paradigma, de acuerdo Crabtree y Miller (1992a: 8; cit. por Valles, 2003), “presenta un conjunto entrelazado de supuestos que conciernen a la realidad (ontología), conocimiento de esa realidad (epistemología), y las formas particulares para conocer esa realidad (metodología) (...) Cada investigador debe decidir qué supuestos son aceptables y apropiados para el tema de interés y luego usar métodos consecuentes con el paradigma seleccionado”. Paradigma es

entonces un conjunto integrado por suposiciones acerca de la naturaleza del mundo social, acerca del carácter de conocimiento que nosotros podemos tener acerca de ese mundo y acerca de qué es importante conocer. Están los paradigmas clásicos, racionales o positivistas por un lado; del otro, los naturalistas, constructivistas o interpretativistas. El primer grupo “asume la existencia de una sola realidad objetiva averiguable a través de los cinco sentidos, sujeta a las leyes universales de la ciencia, y manipulable mediante procesos lógicos”, el segundo grupo asumiría la existencia de “realidades múltiples con diferencias entre ellas que no pueden resolverse a través de procesos racionales o aumentando tamaños muestrales” (Erlandson y otros, 1993: 14; cit. por Valles, 2003: 53).

Así, para Guba y Lincoln (1985, 1994; cit. por Valles, 2003: 49), los paradigmas deben entenderse como sistemas de creencias básicas sobre la naturaleza de la realidad investigada (supuesto ontológico), sobre el modelo de relación entre el investigador y lo investigado (supuesto epistemológico), y sobre el modo en que podemos obtener conocimiento de dicha realidad (supuesto metodológico). Lo metodológico no se trata solo de la elección entre métodos cuantitativos o cualitativos, sino que tiene que ver con los procedimientos metódicos que resultan de las posturas ontológicas y epistemológicas adoptadas. Valles (2003: 51) cita un ejemplo de Guba y Lincoln (1994) al señalar que si se ha partido de una realidad ‘real’ (objetivamente aprehensible) y una separación sujeto-objeto, la preocupación metodológica se centrará en el control experimental de posibles factores explicativos alternativos. Si, en cambio, se ha partido de un *realismo crítico* (“una realidad ‘real’ pero solo imperfectamente y probabilísticamente aprehensible”), y se define la comunicación sujeto-objeto, se practicará (dentro de una lógica experimental revisada) un mayor interés por la utilización de métodos y técnicas cualitativos.

Años después, Guba y Lincoln (2000) agregaron dos conjuntos más de atributos a los paradigmas. El primero se refería ya a las creencias filosóficas básicas relacionadas con la naturaleza del mundo social (ontología), la naturaleza del conocimiento social (epistemología), y la forma en las cuales tales conocimientos pueden ser desarrollados o construidos (metodología). El segundo conjunto incluye “posiciones de paradigma sobre determinadas cuestiones prácticas”: naturaleza del conocimiento, acumulación del conocimiento, criterios de calidad, valores, ética, postura del investigador, y capacitación. Por ejemplo, en el post-positivismo, los valores

son excluidos y su influencia se niega, mientras que en el constructivismo los valores son incluidos y su influencia es vista como formativa. El tercer conjunto de atributos del paradigma es etiquetado como Temas críticos de la época: axiología; adaptación; acción; control; relación a los fundamentos de la verdad y conocimiento; consideraciones prolongadas de criterios de validez y bondad; y voz, reflexividad, representaciones textuales postmodernas (Greene, 2007: 52-53).

El entendimiento de estos paradigmas nos ayuda a comprender la realidad que los evaluadores y clientes de la misma han percibido en determinado momento y el porqué del uso de ciertos métodos de la evaluación en determinadas épocas, o porqué la importancia de los objetivos, la experimentación o la pluralidad.

Al igual que en la investigación, la evaluación tiene lugar dentro de un contexto socio-histórico específico, donde los evaluadores o los clientes toman decisiones que muestran apego a ciertas posturas ideológicas o sociopolíticas, a su educación, a valores personales, o compromisos, lo cual se refleja en los objetivos y estándares de la evaluación, el enfoque y los métodos elegidos. La ciencia social en la mayoría de los países estuvo dominada por el paradigma post-positivista a través de gran parte del siglo XX. La suposición de este paradigma invoca estandarizar métodos y diseños cuantitativos. Durante las tres últimas décadas del siglo XX, los retos del post-positivismo generaron un interés en otros paradigmas filosóficos más acordes con metodologías cualitativas, incluyendo el interpretativismo, varias formas de constructivismo, y fenomenología. La evaluación no fue ajena a esos paradigmas y enfoques.

De acuerdo con Jennifer Greene (2007: 37), en el debate fueron centrales las diferentes suposiciones acerca de la naturaleza del mundo social y la naturaleza del conocimiento social. En el post-positivismo, el mundo social es asumido como real, existe independientemente de nuestro conocimiento de él y supone que funciona más parecido al mundo físico o natural (así como las plantas reaccionan predeciblemente a diferentes condiciones ambientales de humedad, luz y temperatura, así los humanos reaccionan predeciblemente a las diferentes fuerzas y factores en el ambiente externo). En la mayoría de los paradigmas interpretativos, el mundo social es asumido como en construcción, al menos en parte. Los seres humanos, a diferencia de las plantas, actúan con intencionalidad arraigada en los significados que las personas construyen de varios fenómenos hallados en su vida diaria,

mientras interactúan con otros en una variedad de escenarios. Y es esta construcción de significados lo que guía y da forma al comportamiento humano, más que las fuerzas o factores externos.

El método cuantitativo tradicionalmente se asocia a la medición, al explicar, controlar y predecir fenómenos mediante la utilización, principalmente, de métodos experimentales y matemáticos donde al hacer uso de estos los resultados son objetivos y no hace falta apelar al juicio de la persona ni de los destinatarios. Las preguntas de evaluación deben ser dirigidas rigurosamente, por lo tanto el evaluador debe mantenerse independiente y sus valores no deben inferir con lo que evalúa, pues aspiran “al realismo, a la objetividad, a la explicación causal y a la verdad universal” (Greene, 2005). Este método ha estado ligado a la verificación, es decir, a verificar y comprobar teorías. Para ello recolecta y analiza datos para contestar preguntas y probar hipótesis previamente establecidas, y usa la medición y la estadística para determinar los patrones de comportamiento en una población. Asimismo lleva a cabo relaciones entre variables para realizar afirmaciones y recomendaciones precisas. Parte de que hay una realidad que conocer, su finalidad es reportar qué sucede, es decir, los hechos que nos den información concreta de la realidad que podemos explicar y predecir (Hernández, Fernández, Baptista: 2003).

Por su parte, el método cualitativo, contrario a la medición, se basa en un fundamento humanista para entender la realidad social, aboga por métodos naturalistas y prefiere registrar sus hallazgos en el lenguaje de sus sujetos, considera sus palabras vitales pues estas se convierten en los resultados y descubrimientos de la evaluación. Es considerado subjetivo. Los cualitativistas enfatizan la interpretación, el valor y la naturaleza contextual del conocimiento social (Greene, 2005). Este método busca reconstruir y entender la realidad y está más ligado a la tarea de descubrir y generar teorías, por lo cual las preguntas de evaluación pueden (re) surgir durante el estudio, y basa su recolección de datos en descripciones y observaciones. La finalidad de este método es entender el contexto y/o el punto de vista del actor social (Hernández Sampieri, Fernández Collado, Baptista Lucio, 2003).

La información cualitativa puede ser recolectada a través del diseño, pero algunas veces puede aparecer inesperadamente o ser descubierta a través de la investigación exploratoria. Puede implicar un amplio rango de variables del programa, tales como las necesidades de los beneficiarios, cómo y

porqué un programa comenzó, metas y planes, personal y procedimiento, equipo e instalaciones, operaciones y gastos, y resultados pretendidos y no pretendidos (Stufflebeam y Shinkfield, 2007: 597). Recomienda una constante interrelación con políticos, promotores y representantes de los intereses afectados para pactar la manera y forma de realizar el proceso evaluador (Subirats, 1994).

Las primeras evaluaciones suponían que los programas producían ciertos resultados deseados, el propósito de la evaluación consistiría en comprobar y medir esos efectos. Sin embargo, con el paso del tiempo se fue consciente de que los programas producían otros efectos no previstos, el propósito ya no era solo verificar sino también descubrir cómo el programa podría alcanzar los efectos deseados. Estas primeras evaluaciones cuantitativas produjeron en los administradores de los programas ciertas decepciones, pues ellos esperaban que las evaluaciones les hicieran saber, entre otras cosas, qué funcionaba mejor o qué aspectos había que cambiar; sin embargo, eso no ocurría. Al componente sumativo de la evaluación se comenzó también a pedir el componente formativo. En este sentido, el *Joint Committee* (1994) establece que el análisis de la información cuantitativa o cualitativa, además de ser apropiado y sistemático, debe responder efectivamente las preguntas de la evaluación. Estas preguntas de la evaluación y necesidades de los clientes son el centro de la evaluación, para responderlas de la mejor manera, más que la distinción entre diseño cualitativo o cuantitativo se han incrementado los diseños y prácticas de evaluación con métodos mixtos, pudiendo dar mayor énfasis a lo cuantitativo que a lo cualitativo o viceversa dependiendo de la clase y calidad de información que dé validez a las respuestas planteadas en las preguntas evaluativas y responda a las necesidades de su audiencia, logrando así conclusiones justificables acerca del valor del programa evaluado.

En la evaluación educativa, por ejemplo, la historia no ha sido diferente, los interesados en la educación comenzaron a ver la investigación y la evaluación tradicional como una abstracción de la realidad de la vida cotidiana de las escuelas. Fue importante ya no solo saber qué porcentaje de niños en un país no sabía matemáticas, leer o escribir, sino ahora interesaba por qué no sabían y qué es lo que se necesitaba para que pudieran aprenderlo. A todo esto se sumaban las grandes cantidades de dinero que se destinaban a los programas y la creciente complejidad de la sociedad (diferencias en cuanto a cultura, conducta, raza, lo político del tema, etcétera). El paso

de la simple verificación (cuantificación a través del énfasis en exámenes psicométricos a los estudiantes) al descubrimiento (cualitativo) era ya una realidad. La unidad de análisis comenzó a desplazarse de una exclusividad en el individuo, como quien aprende a una atención también en el programa como el que enseña y la escuela como el lugar en donde se lleva a cabo la tarea educativa. El evaluador debía tener claro y admitir que no toda evaluación requiere solo de análisis cuantitativos, y que no es aceptable elegir el método y las técnicas que ellos conozcan y con las que se sientan mejor, al elegirlos se debe tener en cuenta las necesidades de la audiencia y las preguntas centrales de la evaluación, y solo con base en eso seleccionar las que mejor respondan a ellos.

Así entonces, el uso del método exclusivamente cuantitativo ha sido insuficiente. Para un panorama completo en evaluaciones integrales, lo más recomendable es el empleo de ambos métodos a través de diversas técnicas y su correspondiente triangulación. Pues las estadísticas y demás análisis cuantitativo, requieren muchas veces de un complemento y confirmación cualitativo. El papel de ambos análisis es dar las bases para unos juicios finales y recomendaciones válidas y justificadas del programa ante las audiencias interesadas. Hernández López (2001: 136) apunta que la práctica profesional de la evaluación de programas ha revelado cómo las evaluaciones cualitativas suelen restringirse a programas locales, es decir, muy contextualizados, mientras que las cuantitativas son empleadas en aquellos que superan ese ámbito local y son más ambiciosos en sus objetivos de alcance.

Podemos concluir que lo que distingue a ambas posturas metodológicas es su visión respecto de la naturaleza del orden social. El enfoque que de la vida social hace el paradigma cuantitativo, emplea los supuestos mecanicistas y estáticos del modelo positivista de las ciencias naturales. Por otra parte, el paradigma cualitativo posee un fundamento humanista para entender la realidad social de la posición idealista que resalta una concepción evolutiva y negociada del orden social, percibe la vida social como la creatividad compartida de los individuos; el mundo social no es fijo ni estático, es cambiante y dinámico y los individuos son conceptuados como agentes activos en la construcción y determinación de las realidades que encuentran. Este interés por los significados sociales y la insistencia en que tales significados solo pueden ser examinados en el contexto de la interacción de los individuos, es lo que distingue a este paradigma del modelo de investigación de las ciencias naturales (Filstead, 2005: 62-64).

5.1. Método mixto. La combinación de métodos cualitativos y cuantitativos en la evaluación

Como respuesta al debate sobre la mejor técnica utilizada en la evaluación, Jennifer Greene (2005) ha escrito sobre el método mixto (*Mixed-method*) como el método que involucra el uso de dos o más tipos diferentes de recolección de datos y técnicas de análisis en la evaluación o investigación social, convirtiéndose en un método capaz de generar mejor conocimiento. Así, por ejemplo, podemos combinar un experimento y una etnografía o mezclar una encuesta con un estudio de caso. Greene considera esto más factible que utilizar diferentes diseños de investigación dentro de un mismo estudio. Lo que en realidad importa en este método no es el carácter numérico o narrativo de los métodos usados, sino lo que tiene que ver con el grado de estandarización y estructura (por ejemplo, decidir por anticipado las preguntas específicas a realizar o permitir que diferentes preguntas surjan durante la entrevista) o el grado de generalidad o particularidad (extender los hallazgos a otras poblaciones o profundizar sobre un contexto o población en particular).

La investigación del método mixto ofrece una serie de propósitos desarrollados con el paso del tiempo y que no son excluyentes entre sí sino que varios de ellos se pueden conjuntar en la práctica de la evaluación. Así, al llevar a cabo una investigación con el método mixto esta puede ofrecer:

Un entendimiento más defensivo, con más énfasis en la validez y credibilidad (ejemplo, observaciones estructuradas y análisis cualitativos de documentos).

Un entendimiento más extenso, desarrollando descripciones más completas de nuestro entorno social a través del uso de múltiples perspectivas (por ejemplo, observaciones estructuradas y entrevistas abiertas).

Un entendimiento más intuitivo, con nuevas ideas, perspectivas frescas y conceptos creativos.

Un entendimiento con gran valor de conciencia y con una mayor diversidad de valores y posturas a través de la inclusión de diferentes métodos que en sí mismos proponen diferentes valores (por ejemplo, entrevistas abiertas con cuestionarios estructurados).

En este tipo de método, además de mezclarse lo cuantitativo con lo cualitativo, surge el debate acerca de si los investigadores también mezclan supuestos filosóficos y si deberían hacerlo. Esto último implicaría que los supuestos clave incluyen visiones del mundo social (por ejemplo, realismo o construccionismo), perspectivas respecto de la naturaleza del conocimiento social (por ejemplo, objetivas o cargado de valores) y posiciones respecto de qué es más importante conocer (por ejemplo, relaciones causales generalizables o importancia del contexto).

Son tres las posturas que existen: la primera argumenta que los supuestos filosóficos son herramientas conceptuales útiles, pero no deberían guiar las decisiones prácticas, pues el diseño y el método de estas últimas debería dirigirse por las demandas del contexto; la segunda argumenta a favor de la mezcla de supuestos filosóficos, al mismo tiempo que se mezclan los métodos (postura dialéctica), aquí las posibles tensiones provenientes de los diferentes supuestos se toman como generadoras de nuevo conocimiento. La tercera incluye marcos filosóficos dentro de los cuales diversos métodos y supuestos pueden cómodamente encontrarse (postura pragmática).

En la práctica del método mixto, Greene establece que es poco probable que una sola guía prescriptiva pueda capturar la gran gama de combinaciones y facetas de su posible diseño, análisis, calidad e informe. Lo que más bien encontramos son tipologías, posturas y dimensiones a tener en cuenta. Respecto del diseño, hay tres dimensiones claves. La primera se refiere a la integración o separación de los métodos durante el estudio y al punto en que las conclusiones y deducciones son comparadas y conectadas. Así, por ejemplo, en un diseño integrado, los datos de los diferentes métodos pueden informar del diseño de un instrumento en particular u ofrecernos un diseño muestral de otro, combinándose al final los diferentes tipos de datos en el análisis, produciendo un solo conjunto de resultados y deducciones, de tal forma que no se muestran más como distintos. Por su parte, en un diseño de componentes, los datos conservan su forma original durante todo el proceso, y las conclusiones y deducciones buscan una armonía y conexión más que una integración o combinación completa. La segunda dimensión se refiere a la importancia de los diferentes métodos involucrados, si una metodología es más dominante, por lo general el investigador se apegará más a su guía tradicional, pero el reto en la práctica del método mixto radica en otorgar una relativa igualdad de importancia a los diferentes métodos. La tercera dimensión se refiere a la implementación de los diferentes métodos,

la cual puede ser de manera simultánea o secuencial, ya sea por razones conceptuales o prácticas.

En cuanto al análisis de datos, las ideas no están muy catalogadas, y menos aún para los diseños integrados, pero de las investigaciones realizadas es posible identificar diferentes aproximaciones, una sugiere un camino paralelo para los diseños de componentes, un camino cruzado para los dos diseños o un solo camino para los diseños integrados. En el análisis de caminos cruzados, los análisis son iniciados de manera separada, cada uno desde su respectivo ámbito, ya sea cualitativo o cuantitativo, donde después uno de esos datos puede ser transformado y después cruzado con el otro para su comparación y posterior análisis. En el análisis de un solo camino, los datos de los diversos métodos son fusionados en una sola corriente. Otro de los retos durante la práctica se refiere al desarrollo. Los criterios y procedimientos para la calidad se relacionan con las posturas de las diferentes tradiciones metodológicas, con los diferentes métodos de relativa igual importancia y el esfuerzo de su integración. Por último, algunos investigadores experimentan con el uso de formas de representación mixta para reportar los resultados del método mixto. Este método aún sigue desarrollándose, pero proporciona una opción para conjugar ambos tipos de técnicas con el rigor que una evaluación requiere.

6. Los enfoques evaluativos más relevantes para inicios del presente siglo

Cada década ha marcado diversas corrientes en evaluación, algunos evaluadores preferían utilizar los objetivos declarados en el programa para definir los límites de su estudio mientras otros optaban por utilizar los valores de los actores implicados. Desde las primeras prácticas evaluativas, varios evaluadores han propuesto su propia metodología para desarrollar una evaluación, a esta manera de priorizar un aspecto, de desarrollar y llevar a cabo una evaluación, es a lo que se llama enfoques o modelos evaluativos. Más concretamente, los enfoques de evaluación ofrecen un esquema o diseño general que ha de llevarse a cabo, los distintos criterios de valor para juzgar el objeto evaluado, y las técnicas y procedimientos para recopilar la información y realizar el respectivo análisis de una determinada realidad. Estos enfoques no son de ninguna manera excluyentes entre sí, sino complementarios y, como bien lo menciona Vedung (1997: 120), es importante tener en cuenta que todo modelo de evaluación ofrece solo perspectivas y respuestas parciales.

Desde 1960, comenzó a surgir una gran variedad de enfoques evaluativos, sobre todo en Estados Unidos, y al día de hoy se encuentra en la literatura una amplia diversidad de ellos. Una forma típica de clasificarlos ha sido dividirlos en enfoques clásicos y alternativos. Los primeros con tendencias más racionales en cuanto a su metodología y resultados, dejando de lado las condiciones externas a los programas, y dando por sentado aspectos que influyen de manera importante en la evaluación y sus resultados. Las

limitaciones de estas evaluaciones se tradujeron en la irrelevancia de sus resultados, debido a que la mayoría de las veces la evaluación no respondía a las necesidades de los decisores o grupos sociales, quienes tenían muy poca participación en el proceso evaluativo. Como respuesta, aparecieron enfoques alternativos preocupados más por incluir a la gran mayoría de los agentes involucrados en los programas y su evaluación.

Al igual que Stufflebeam (2007), se considera que la importancia de analizar los distintos enfoques evaluativos reside en que permite tanto a evaluadores como a clientes diferenciar los estudios meramente evaluativos de aquellos que en nombre de la evaluación son estudios que se prestan a la corrupción o responden a intereses meramente políticos. Desde la perspectiva teórica, el análisis de los enfoques evaluativos ayuda a identificar, examinar y establecer aspectos conceptuales y técnicos para el desarrollo de la evaluación como disciplina. En un sentido más operativo y práctico, ayuda a los evaluadores a considerar y aplicar selectivamente, bajo ciertas circunstancias, marcos de trabajo apropiados para cada evaluación. También, al analizar los distintos enfoques evaluativos es posible ubicar sus fortalezas y debilidades, y determinar el mérito en el uso de alguno de estos, además del cuándo y cómo, es más adecuado aplicarlos e incluso poder hacer mejoras y elaborar mejores alternativas, y según sea el caso, conceptualizar enfoques híbridos para la evaluación de programas.

Stake (2006) ha hecho una relación de los enfoques con base en los roles habituales de la evaluación. Dentro de los enfoques que tienen una predisposición a la medición de la consecución de los objetivos, se encuentran: la evaluación instrumentalista, la evaluación de la discrepancia, la evaluación de modelos lógicos, la evaluación de resultados, la evaluación de impactos, la evaluación intencional y la evaluación de ecuaciones de regresiones. Los enfoques que se inclinan por el desarrollo organizativo incluyen: la evaluación basada en las decisiones, la evaluación basada en la utilización, la evaluación centrada en el cliente, el autoestudio organizativo, la investigación-acción, la evaluación emancipadora y la acreditación institucional. Entre los enfoques de evaluación empleados por los evaluadores con una inclinación a la detección de la calidad holística, están: la evaluación naturalista, la evaluación interpretativa, la evaluación constructivista, la evaluación comprensiva y la evaluación transaccional. Los enfoques evaluadores que muestran una predisposición a la búsqueda y a la valoración de la calidad de la estrategia o política, son: la evaluación

basada en la teoría, los estudios experimentales, la evaluación basada en las ciencias sociales y la evaluación con modelos causales. Finalmente, los enfoques evaluadores con una inclinación por el estudio de las cuestiones humanas y por la contribución a la acción social, son: la evaluación participativa, la evaluación colaborativa, la evaluación democrática, la evaluación emancipadora y la evaluación personalista.

Stufflebeam (2000, 2007), por su parte, ha realizado una clasificación de los enfoques evaluativos, partiendo de los enfoques más representativos del siglo XX y ha obtenido una lista de los enfoques que desde su punto de vista indican su validez en este siglo XXI. Para esta clasificación se ha basado en el grado de conformidad a la definición de evaluación que la considera como un “estudio diseñado y dirigido a ayudar a cierta audiencia a valorar el mérito y valor de un objeto”. Usando esta definición, en un primer estudio del año 2000, los enfoques de evaluación de programas fueron clasificados por Stufflebeam en cuatro categorías, para 2007 se agregó una categoría más:

1. Pseudoevaluaciones. Incluye los enfoques que promueven hallazgos inválidos o incompletos y que generalmente están motivados por objetivos políticos. Aquí encontramos al enfoque de relaciones públicas y los estudios políticamente controlados. Para 2007, Stufflebeam agregó tres enfoques más: las evaluaciones complacientes, evaluación por pretexto (*evaluation by pretext*) y empoderamiento bajo la apariencia de evaluación.
2. Enfoques orientados a las preguntas y métodos de la evaluación (Cuasievaluaciones). Por una parte, se centra en preguntas y respuestas que pueden ser o no suficientes para valorar el mérito o valor de un programa. Por otra parte, hace uso de algunos métodos preferentes. Estos enfoques pueden utilizar como sus puntos de partida objetivos operacionales, dispositivos de medida estandarizados, procesos de costo-análisis, juicios de expertos, una teoría o modelo de algún programa, procedimientos de estudios de caso o emplear una mezcla de métodos cualitativos y cuantitativos. La mayoría enfatiza la técnica cualitativa y postula que es mejor responder a pocas preguntas que intentar una valoración extensa del mérito o valor de alguna cosa. Se les llama cuasievaluaciones porque algunas veces dan evidencia que permite la valoración

completa de un programa, mientras que otras su foco de atención es muy estrecho. Aquí se incluyen a los estudios basados en objetivos; la rendición de cuentas, particularmente el pago por estudios de resultados; los programas de medición de objetivos; la evaluación de resultados como valoración del valor añadido; el meta análisis; los estudios experimentales; la gestión de sistemas de información; el análisis costo beneficio; la aclaración a la audiencia (*clarification hearing*); las evaluaciones de estudios de caso; la crítica y conocimiento; la evaluación basada en la teoría del programa y los estudios de métodos mixtos. Para 2007 se agregó también el método del caso de éxito (*the success case method*).

3. Enfoques orientados a la mejora y a la rendición de cuentas. Pone énfasis en la necesidad de valorar de una manera completa el mérito y valor de un programa. Busca la amplitud al considerar todas las preguntas y criterios necesarios para valorar el programa, y suele utilizar las necesidades de valoración de las distintas partes interesadas en el programa, como su criterio base para valorar el mérito o valor del mismo. Al mismo tiempo, buscan examinar de manera completa la pertinencia de los criterios técnicos y económicos para juzgar el programa y buscan todos los resultados relevantes y no solo los estipulados en los objetivos del programa. En cuanto a los métodos, utiliza tanto los cuantitativos como cualitativos. Hablamos de enfoques orientados a la decisión y a la rendición de cuentas, los orientados al consumidor y los enfoques de acreditación y certificación.
4. Enfoques dirigidos a la agenda social y al apoyo. Estos enfoques pretenden hacer una diferencia en la sociedad, a través de la evaluación de programas. Buscan asegurar que todos los sectores de la sociedad tengan igual acceso a los beneficios sociales. La perspectiva que emplean para describir, investigar y juzgar el programa es tanto la de las partes interesadas como la de los expertos. En estos enfoques encontramos a los estudios centrados en el cliente o la evaluación receptiva, la evaluación constructivista y la evaluación democrática deliberativa.
5. Enfoques eclécticos. Este ha sido el quinto enfoque que Stufflebeam agregó a su clasificación. Aquí los evaluadores no se comprometen con ninguna filosofía en particular, ni con un enfoque metodológico

o misión social. En vez de eso, han avanzado en enfoques pragmáticos que utilizan un amplio rango de ideas y procedimientos de otros enfoques evaluativos. Este tipo de enfoques no tiene lealtad a ninguna escuela de pensamiento evaluativo, pero selecciona de estas las doctrinas y las aplica a sus estudios. Sus enfoques son diseñados para dar cabida a las necesidades y preferencias de un amplio rango de clientes de la evaluación y de tareas evaluativas. En consecuencia, este tipo de evaluadores emplea, ya sea bases filosóficas, marcos conceptuales, y procedimientos que los puedan conducir a los objetivos específicos de la evaluación y cumplir los deseos de los clientes de dicha evaluación. Muy probablemente ellos emplearán selectivamente elementos de varios enfoques evaluativos. El enfoque que Stufflebeam considera en esta categoría es la evaluación centrada en la utilización, enfoque que en su anterior clasificación del año 2000, incluía en los enfoques dirigidos a la agenda social y al apoyo.

Después de analizar todos los enfoques, Stufflebeam concluyó, en primera instancia, que solo nueve de ellos podían ser identificados como los más fuertes y prometedores para continuar su uso y desarrollo en este siglo¹³. Estos enfoques eran, con base en un mayor mérito y dentro de las categoría de mejoramiento y rendición de cuentas, agenda social y apoyo, y enfoques de evaluación orientados a las preguntas y métodos: los estudios orientados a la decisión o rendición de cuentas, los estudios enfocados al consumidor, los enfoques de acreditación o certificación, la evaluación centrada en la utilización, los estudios centrados en el cliente, la evaluación democrática deliberativa, la evaluación constructivista, el estudio de caso y la evaluación de resultados como valoración del valor añadido. Pero años después, redujo su lista a ocho enfoques, donde descartó a la evaluación de resultados como valoración del valor añadido, los enfoques de acreditación o certificación y la evaluación democrática deliberativa. En su lugar agregó la evaluación basada en objetivos y el diseño experimental, pues se cree que pueden ser buenas alternativas a la evaluación. Sin embargo,

13 Para llegar a la clasificación de sus enfoques (ordenados con base en el mérito) y después a la elección de los enfoques, Stufflebeam se basó en los “Estándares de Evaluación de Programas” del *Joint Committee* (utilidad, factibilidad, legitimidad y adecuación) y después dedujo los enfoques utilizando una lista de control especial de Estándares.

se sigue creyendo que los otros seis enfoques son más recomendados para considerarse en una evaluación de programas.

La clasificación de evaluación presentada enseguida, está basada en esta propuesta de Stufflebeam y en los enfoques relevantes para este siglo. Atendemos a esta clasificación debido a que parte de una definición de evaluación ampliamente aceptada y que también se comparte: “valorar el mérito y valor de un objeto”, además de que los enfoques son los más relevantes y citados en la literatura de evaluación. Se considera oportuno dividir los enfoques en clásicos y alternativos para el siglo XXI, donde ya la última clasificación de Stufflebeam recoge los enfoques clásicos. A continuación, se presenta una breve exposición de los enfoques. En esta se incluye un pequeño análisis de cada uno de estos, mencionando su principal exponente, sus principales características, su propósito, las principales preguntas a las cuales responde, los valores y el método que utiliza, así como sus ventajas y desventajas. Los enfoques analizados se presentan en la tabla 2.

Tabla 2. Enfoques clásicos vs enfoques alternativos para el siglo XXI

Enfoques clásicos	Enfoques alternativos para el siglo XXI
Enfoque por objetivos.	Estudios orientados a la decisión y a la rendición de cuentas.
Modelos experimentalistas.	Estudios orientados al consumidor.
	Enfoques de acreditación y certificación.
	Evaluación centrada en la utilización.
	Estudios centrados en el cliente.
	Evaluación democrática deliberativa.
	Evaluación constructivista.
	Estudio de casos.
	Evaluación de resultados como valoración del valor añadido.

Fuente: Stufflebeam 2000, 2007.

6.1. Evaluación por objetivos

Modelo propuesto por Tyler (1942), usado en la realización de las primeras evaluaciones. Él definió a la evaluación como “el proceso de determinar en qué medida los objetivos de un programa se habían alcanzado” (Tyler, 1949). El propósito, por lo tanto, era determinar si los objetivos del programa habían sido alcanzados o no. Para ello se tomaba como punto de partida las metas y objetivos del programa (tal cual aparecían enunciados), para después delimitarlos de manera jerárquica y medible, y proceder a la selección y puesta en práctica de los instrumentos adecuados para medir las escenas del programa en que se realizaban o no los objetivos. El proceso finalizaba con el análisis comparativo de la información recopilada y de la que se quería lograr, averiguando hasta qué punto el programa había favorecido o dificultado la consecución de los objetivos, deduciéndose así si se había tenido éxito en el trabajo valorado. Los métodos utilizados son principalmente los cuantitativos, ya que tienden a ser sistemáticos y racionales, los test son particularmente usados.

La evaluación por objetivos tiende a emitir juicios sobre el producto final. Se basa en el supuesto de que es posible dar respuestas científicas u objetivas a las preguntas del proceso de evaluación, por lo que otorga un papel secundario al consenso y la relación con los actores del programa. Su ventaja reside en que al ser común su uso, los administradores de los programas ya tienen experiencia en ella. Sus desventajas radican en que pareciera sencillo de realizar, pero en la práctica surgen problemas debido a que no siempre los objetivos se presentan de una manera clara y operacional, pues muchas veces están especificados en términos generales, abstractos y confusos. Además, al centrarse en una información concreta y final otorgan pocas posibilidades para la mejora del programa y, a menudo, la información que ofrecen es limitada y no proporciona las suficientes bases para juzgar el mérito o valor del objeto. Una desventaja más es que no revelan los efectos secundarios tanto positivos como negativos.

6.2. Modelos experimentalistas

En 1966, D. Campbell y J. Stanley introducen los diseños preexperimentales, experimentales y cuasiexperimentales, que han constituido una de las formas tradicionales de evaluación. Estos diseños consideran implícitamente a la acción pública como una ciencia en la que se pueden realizar

experimentos sobre la realidad. El propósito es determinar las relaciones causales entre variables dependientes e independientes específicas (por ejemplo, entre un método educativo y el rendimiento de los estudiantes en pruebas estandarizadas), basándose en “cinco principios fundamentales: las relaciones causales, comparación de medidas, asignación aleatoria, representatividad y posibilidad de replicar el experimento” (Monnier, 1995: 44).

El principio que siguen los experimentalistas es la formulación de una hipótesis donde una acción X producirá un cambio Y en la dirección pretendida, y para esto es necesario no introducir modificaciones al tratamiento experimental mientras dure el programa. Los resultados se basan en la obtención de datos cuantitativos y el método que utilizan es la comparación de los resultados alcanzados por un grupo sometido al programa (grupo experimental) y otro fuera del mismo (grupo de control). Lo anterior implica que de la población a la que va dirigido el programa se eligen unidades (grupos, comunidades, escuelas, aulas...) que recibirán la atención-beneficio del mismo y, a su vez, existe otro grupo que no recibirá la atención, pero servirá para determinar los efectos de la acción. Un punto importante es la asignación aleatoria de los individuos a cada uno de los dos grupos, ya que “únicamente una muestra aleatoria asegura que todos los individuos de una misma población tengan la misma oportunidad de ser seleccionados y por tanto que los resultados puedan ser generalizados a toda la población” (Langbein, 1980: 65). El éxito en el programa radica atribuir al grupo experimental mejores resultados que al grupo de control.

Las preguntas de evaluación son establecidas por los investigadores, los diseñadores del programa o los tomadores de decisiones. La pregunta más frecuente tiene que ver con el conocer cuáles son los efectos de una intervención en una variable específica de resultados. Una ventaja de este enfoque es que proporciona métodos sólidos para establecer relaciones causales inequívocas entre el tratamiento y las variables de resultado. Sin embargo, a pesar de que su uso y éxito en experimentos dentro de campos como la medicina y la agricultura les ha dado credibilidad, estos modelos resultan difíciles de aplicar en la práctica social debido a que en los programas sociales influyen muchos factores externos que no pueden controlarse dentro de un “laboratorio” (cambios en el contexto, comportamiento de los diversos actores involucrados...), esto ha provocado que hoy en día este método resulte no muy positivo para realizarse en el ámbito social.

6.3. Estudios orientados a la decisión y a la rendición de cuentas

La parte orientada a la decisión de este enfoque, hace énfasis en el uso productivo de la evaluación para mejorar el programa; por su parte, la rendición de cuentas tiene un uso retroactivo al juzgar el mérito y el valor del programa. El propósito más importante de la evaluación no es demostrar, sino mejorar, y por tal razón pretende proporcionar a los responsables de tomar decisiones información adecuada y bases valorativas para actuar y justificar decisiones que resulten en un uso informado del desarrollo, entrega y toma de decisiones del costo efectividad de los servicios. Para esto, se involucra tanto a los responsables de tomar decisiones como a las distintas partes interesadas ¹⁴ de todos los niveles organizativos del programa para establecer el diseño y desarrollo de la evaluación. De ambos actores se toman en cuenta sus preguntas más importantes y se les entregan los resultados de una manera puntual.

Este intercambio entre el evaluador y los diversos actores supone, por una parte, que la información producida sea relevante y, por otra, que ayude a determinar qué reportes de evaluación serían los más útiles para el personal del programa y demás partes interesadas. Este intercambio entre el evaluador y los diversos actores, además de la información producida, sirven de vehículo para una retroalimentación de la evaluación usada en la mejora del programa. Esa misma información ayuda también a producir un expediente de rendición de cuentas. Mediante la implicación de las distintas partes interesadas, a través de los procesos de evaluación, los evaluadores sientan las bases para aportarles el entendimiento y valor de los procesos de evaluación y así aplicar los hallazgos.

Este enfoque proporciona una evaluación tanto formativa como sumativa al centrar sus necesidades de información en la valoración de necesidades, problemas, y oportunidades; identificación y valoración de programas alternativos; valoración de planes de programas; valoración de las aptitudes y desempeño del personal; valoración de las instalaciones y materiales del programa; monitoreo y valoración del proceso; valoración de los resultados pretendidos y no pretendidos, de corto y largo alcance,

14 Aquí se puede incluir a todas las personas y grupos que deban hacer elecciones relacionadas con el comienzo, planeación, implementación y uso de los servicios de un programa.

y valoración de costo-efectividad. El enfoque logra un balance en el uso de los métodos cualitativos y cuantitativos. Utiliza encuestas, necesidades valorativas, estudios de caso, equipos de apoyo, observaciones, entrevistas y diseños cuasiexperimentales y experimentales.

Entre los representantes de este enfoque se encuentra Cronbach (1963), quien sin utilizar los términos de evaluación formativa y sumativa los definió, y recomendó a los educadores reorientar sus evaluaciones y pasar de una orientación centrada en los objetivos a una preocupación por tomar mejores decisiones en los programas.

Stufflebeam (1966) es otro exponente. Él argumentó que las evaluaciones deberían ayudar al personal del programa a tomar y defender las decisiones clave que reúnan las necesidades de los beneficiarios, al mismo tiempo apoyaba una orientación a la mejora de la evaluación, también enfatizó que los evaluadores debían proporcionar información útil para la toma de decisiones y facilitar información para la rendición de cuentas. Definió a la evaluación como “el proceso de identificar, obtener y proporcionar información útil y descriptiva acerca del valor y el mérito de las metas, la planificación, la realización y el impacto de un objeto determinado, con el fin de servir de guía para la toma de decisiones, solucionar los problemas de responsabilidad y promover la comprensión de los fenómenos implicados” (Stufflebeam, 1987: 183). Esta definición refleja su propuesta de evaluación y resume los conceptos clave de su modelo planteado: CIPP (Contexto, Insumos, Proceso, Producto), cuyo propósito principal es generar información útil. Este modelo se basa en cuatro tipos de evaluación:

1. Evaluación del contexto. Caracteriza el marco en el que se desenvuelve el programa, identifica sus virtudes y defectos, e implica un análisis y valoración de necesidades para establecer objetivos coherentes con dichas necesidades. En esta etapa se define el contexto institucional, se identifica la población objetivo y se valoran sus necesidades y diagnostican sus problemas. Este tipo de evaluación ayudará al establecimiento o ajuste (según sea el caso) de las metas.
2. Evaluación de los insumos (entrada). Consiste en el análisis del programa y la planificación de la intervención, así como en la identificación y valoración de los métodos aplicables. Su propósito es ayudar a establecer alternativas adecuadas a las necesidades del

contexto y desarrollar un plan que sirva a esos propósitos, así como detectar aquellas propuestas innovadoras que fracasarían en caso de ser implantadas. Las propuestas deben ser valoradas de acuerdo con criterios como la contribución e interés en las necesidades, los problemas y los objetivos establecidos, su solvencia y viabilidad.

3. Evaluación del proceso. Es una comprobación continua de la realización del plan. Analiza la ejecución de la intervención, las actividades desarrolladas y la implementación del programa para presentar un informe del mismo, comparándolo con lo que se había planificado. Su misión es ayudar a que el programa se lleve a cabo tal y como se ha planteado o, en su caso, modificarlo.
4. Evaluación del producto. Su propósito es valorar, interpretar y juzgar los logros de un programa. Consiste en el análisis de los resultados con base en los objetivos y la información obtenida en las otras etapas, conociendo hasta qué punto el programa ha satisfecho las necesidades del grupo al cual pretendía servir, así como los efectos, deseados o no, del programa.

Una de las ventajas de este enfoque es que incita a todos los involucrados en el programa a usar la evaluación de manera continua, a planear sistemáticamente y a implementar programas que reúnan las necesidades de la población objetivo. Ayuda también a tomar decisiones a todos los niveles del programa y hacer énfasis en la mejora. De igual manera, presenta información y un fundamento para ayudar al personal del programa a rendir cuentas por sus decisiones y acciones en el mismo. El enfoque involucra a todas las partes interesadas en el proceso de evaluación, para asegurar que sus necesidades de evaluación estén bien establecidas, y para alentarlos y apoyarlos a hacer un uso efectivo de los hallazgos evaluativos. Es una evaluación completa, ya que atiende el contexto, la entrada (*input*), el proceso y los resultados (*outcomes*).

Su mayor desventaja es que la colaboración requerida entre un evaluador y las distintas partes interesadas podría impedir la evaluación en ciertas etapas y/o influenciar sus resultados, especialmente cuando la situación evaluativa está cargada políticamente. Otra desventaja podría darse cuando los evaluadores están influenciando activamente el curso de un programa y se identifican muy estrechamente con él, de tal manera que pierden algo de independencia (necesaria para proveer reportes objetivos).

6.4. Estudios orientados al consumidor

El enfoque considera el bienestar del consumidor como la justificación principal de un programa. El propósito es juzgar el mérito y valor de los productos y servicios de distintos programas y así ayudar a los contribuyentes, profesionales y beneficiarios potenciales a hacer elecciones acertadas. Las preguntas para este enfoque son derivadas de la sociedad, de los componentes del programa, y especialmente del marco de referencia del evaluador. La pregunta general es: ¿cuál de las diferentes alternativas de programas es la mejor elección, dados sus costos diferenciales, las necesidades del grupo consumidor, los valores de la sociedad en general y la evidencia tanto de resultados positivos como negativos? Aquí, el evaluador debería ayudar a los diseñadores a producir y entregar servicios y productos que sean de una excelente calidad y de uso para el consumidor (por ejemplo, estudiantes, padres y madres de familia, docentes y contribuyentes). Así, el evaluador debería ayudar a los consumidores a identificar y valorar el mérito y valor de los programas, servicios y productos ofrecidos.

Los temas valorativos son muy amplios e incluyen la descripción del programa, los antecedentes y el contexto, los clientes, los consumidores, los recursos, las funciones, el sistema de entrega, los valores, los estándares, el proceso, los resultados, los costes, los opositores, la generalización, la trascendencia estadística, las necesidades valorativas, la importancia práctica, las recomendaciones, los reportes, y la metaevaluación. El proceso de evaluación comienza con la consideración de un amplio rango de estos temas, después compila información de todos ellos y finalmente termina en un juicio compacto del mérito y valor del programa. Los métodos que utiliza son las listas de control, necesidades valorativas, evaluación libre de objetivos, diseños experimentales y cuasiexperimentales, análisis del *modus operandi*, aplicación de códigos de conducta ética y análisis de costes (Scriven, 1972). El enfoque se centra en emplear una lista de control válida para los aspectos clave del programa. Scriven (1991) desarrolló una “Lista de verificación de Evaluación” (*Key Evaluation Checklist*) para este propósito.

Asimismo, Scriven (1967, 1973) fue un pionero en la aplicación de este enfoque con su evaluación no centrada en objetivos o libre de objetivos, conocido en inglés como *goal-free evaluation*, donde no toma como criterio de evaluación los objetivos del programa sino más bien las necesidades

fundamentales de los usuarios-consumidores de un programa. Sostiene que el evaluador no solo no debe basar su evaluación en los objetivos del programa, sino debe procurar deliberadamente no informarse sobre ellos con el propósito de no caer en sesgos tendenciosos, por lo tanto, debe mantenerse independiente y no dejarse manejar por los intereses de los involucrados. Para Scriven, el hecho de que los objetivos estipulados en el programa no se cumplan no resulta un problema si en cambio el efecto fue beneficioso para la población a pesar de que este no haya sido intencionado o que, por el contrario, los objetivos se logren mediante efectos indeseables. A Scriven también se le deben los términos de evaluación formativa y sumativa. De acuerdo con él, la meta fundamental de una evaluación válida es juzgar el mérito del programa, valorarlo comparativamente y valorarlo en conjunto. Se espera también que los evaluadores sometan sus evaluaciones de programas a evaluaciones, y a eso lo llamó metaevaluación.

Una de las ventajas de la evaluación orientada al consumidor es que es un intento valorativo para proteger a los consumidores de programas, servicios y productos de mala calidad, y guiarlos a la defensa y uso de aquellas contribuciones que mejor y más efectivamente establezcan sus necesidades. El enfoque hace énfasis en valoraciones independientes y objetivas, y su esfuerzo reside en lograr una valoración integral del mérito y valor para producir una alta credibilidad con el grupo de consumidores. Su principal desventaja viene de esta independencia de los evaluadores, quienes pueden no ayudar a hacer un mejor trabajo en servir a los consumidores.

6.5. Enfoques de acreditación y certificación

Actualmente, muchas instituciones educativas, hospitales y otras organizaciones de servicio, son sujeto de un estudio de acreditación, y muchos profesionales, en una u otra ocasión, se han encontrado con los requerimientos de la certificación para lograr una posición. Así, el propósito de este enfoque es determinar si las instituciones, programas institucionales, y/o el personal deberían ser aprobados u homologados para realizar funciones específicas, es decir, probar si reúnen los requisitos de las profesiones y áreas de servicio, o si son apropiados para servir a las funciones que la sociedad les ha designado. Las preguntas en este estudio surgen del cuerpo de acreditación o certificación, y la pregunta que se establece es si

los programas y personal de las instituciones reúnen los estándares mínimos, y cómo puede ser mejorada su actuación. Así, este enfoque tiene un matiz para la rendición de cuentas con elementos de mejora.

Los valores utilizados en este tipo de estudio son usualmente los lineamientos y criterios que algún cuerpo de certificación y acreditación ha adoptado. Los métodos comúnmente usados son el auto-estudio y el auto-reporte, de manera individual o por institución. En el caso de las instituciones, se les asignan paneles de expertos para que las visiten, verifiquen su auto-reporte y junten la información adicional necesaria. Las bases para el auto-estudio y las visitas por los paneles de expertos son usualmente los lineamientos y criterios que han sido especificados por la agencia de acreditación. En la práctica, existen muchos cuerpos o agencias de acreditación y certificación en diversas áreas y profesiones, como educación o salud, por mencionar algunas.

La mayor ventaja de este tipo de estudios es que ayuda a las personas a realizar juicios informados acerca de la calidad de las organizaciones y programas, y las aptitudes del personal. Las mayores dificultades son que los lineamientos de los cuerpos de acreditación y certificación a menudo ponen énfasis en las entradas y el proceso y no en los resultados. Por otro lado, los auto-estudios y los procesos de visita usados en la acreditación ofrecen muchas oportunidades para la corrupción y la actuación ineficaz, por lo tanto, aquí también se recomienda que los procesos de certificación y acreditación se sometan a metaevaluaciones independientes.

6.6. Evaluación centrada en la utilización

La evaluación centrada en la utilización es un enfoque propuesto por Patton (1980, 1982, 1997), aunque también el Comité Conjunto de Normas para la Evaluación Educativa (*Joint Committee on Standards for Educational Evaluation*) ha hecho énfasis en la utilización de la evaluación (1981, 1994). El propósito principal de este enfoque es que los hallazgos sean utilizados. Se basa en el principio de que es más probable que los sujetos entiendan, valoren, y usen los hallazgos de una evaluación si ellos fueron de manera significativa involucrados en el proyecto. Así pues, el enfoque consiste en realizar un estudio de evaluación en colaboración con un grupo identificado de usuarios prioritarios, con el propósito de centrarse en sus usos pretendidos de la evaluación. Los usuarios prioritarios son elegidos

a través de un análisis cuidadoso y minucioso de las partes interesadas, donde el evaluador identifica sus múltiples perspectivas e intereses que deberían ser representados en el estudio. El evaluador selecciona un grupo que represente a las partes interesadas del programa y los involucra para establecer sus necesidades de evaluación, el uso pretendido de sus hallazgos, la conducción de la evaluación, y el tipo de reportes (ejemplo, oral o escrito) que deberían ser proporcionados. De esta manera, son los usuarios objetivo quienes determinan qué información es requerida, cómo y cuándo debe ser reportada, y cómo será usada. El evaluador ayuda a los usuarios a tomar estas decisiones ofreciéndoles un menú de posibles usos, información, y reportes de la evaluación. Así, dependiendo de las circunstancias, el evaluador puede jugar varios roles: capacitador, experto en medición, experto externo, analista, portavoz o mediador, sin dejar de lado su faceta como negociador.

Todos los aspectos elegidos y aplicados en esta evaluación son para ayudar a los usuarios objetivo a obtener y aplicar los hallazgos de la evaluación en los usos pretendidos. Las bases para interpretar los hallazgos de la evaluación son los valores de los usuarios. Para incrementar la posibilidad de uso de los hallazgos se faculta al grupo representativo de actores con intereses para determinar las preguntas de la evaluación y las necesidades de información. Las preguntas han de ser determinadas localmente, pueden establecer un amplio rango de inquietudes, y probablemente cambiarán conforme la evaluación se desarrolle. Estas preguntas darán las bases para la recolección de la información y reporte de planes y actividades (siempre y cuando los usuarios continúen con su atención en ellas). Los principales posibles usos de los hallazgos de la evaluación contemplados en este enfoque son la valoración del mérito y del valor, la mejora y la generación de conocimiento.

Los métodos y técnicas utilizadas serán aquellas que mejor se adapten a la evaluación a realizar, por ejemplo, cuantitativo y cualitativo, formativo y sumativo, naturalista y experimental. El método de la evaluación de programas centrados en la utilización es etiquetado como “activo-reactivo-adaptativo y sensible”, enfatizando que la metodología se convierte en respuesta a las deliberaciones en curso entre el evaluador y el grupo de clientes, y teniendo en consideración las dinámicas contextuales. Patton (1997: 383) dice que “los evaluadores son activos al presentar a los usuarios potenciales sus mejores juicios acerca del foco de la evaluación

y los métodos; son reactivos al escuchar atenta y respetuosamente las inquietudes de los demás; y son adaptables en encontrar formas para diseñar evaluaciones que incorporen diversos intereses... mientras que reúnen altos estándares de la práctica profesional”.

La mayor ventaja de este enfoque es que involucra a las partes interesadas para determinar los propósitos y procedimientos de la evaluación, y usa su participación para promover el uso de sus hallazgos. Más que tratar de alcanzar y trabajar con todas las partes interesadas, el enfoque de Patton trabaja concretamente con un grupo selecto y representativo de usuarios. Además, atiende de manera cercana la dinámica contextual y pone atención en la necesidad de reunir todos los estándares relevantes de evaluación. Las limitaciones tienen que ver con el reemplazo de los usuarios, ya que esto puede requerir que la evaluación del programa sea renegociada para renovar las perspectivas de los impactos de la evaluación, ocasionando un desvío o retraso en el proceso. Además, el enfoque parece ser vulnerable a la corrupción por los grupos usuarios, a partir de que a ellos se les da mucho control sobre lo que será investigado, las preguntas establecidas, los métodos empleados y la información obtenida. Las partes interesadas con sus intereses en conflicto pueden influenciar de manera inapropiada la evaluación. Este enfoque requiere de una fuerte habilidad para la negociación, comunicación y habilidades políticas por parte del evaluador, así como experiencia en el amplio rango de los métodos de evaluación cuantitativa y cualitativa.

6.7. Estudios centrados en el cliente (o evaluación sensible)

Robert Stake (1983) es su principal exponente, utiliza como principal criterio de valor las necesidades de los clientes de la evaluación con el fin de proporcionarles información evaluativa útil. Para esto, el evaluador debe trabajar con y para apoyar a un grupo diverso de clientes, incluyendo, por ejemplo, maestros, administradores, diseñadores, contribuyentes, legisladores y patrocinadores financieros. Ellos son los clientes en el sentido de que apoyan, desarrollan, administran, u operan directamente los programas bajo estudio y buscan o necesitan el consejo de los evaluadores y un asesoramiento en el entendimiento, juicio y mejora de los programas. El evaluador debe interactuar continuamente con, y responder a, las necesidades evaluativas de los diversos clientes. Para esto, el evaluador suele negociar

con el cliente sobre lo que ha de hacerse y “responde” a lo que quieren saber los distintos destinatarios. De manera simultánea a esto, el evaluador procura obtener información sobre los propósitos, historia, transacciones, y resultados del programa. Se da una especial atención a los efectos secundarios y a los estándares y juicios de varias personas sobre el programa (es obligación del evaluador reflejar todos los intereses pertinentes en la evaluación). Así entonces, el proceso comienza con un análisis de la base conceptual del programa; continúa con la identificación de las normas de calidad establecidas con criterios explícitos por grupos de referencia importantes en cada caso; para proseguir con la descripción detallada del programa que incluye antecedentes, actividades y resultados/consecuencias; finalizando con la valoración en función de comparar los datos descriptivos con las normas de calidad.

La comunidad, los profesionales, y los grupos beneficiarios en el ambiente local, junto con los expertos externos del programa, son quienes establecen las preguntas. En general, quieren conocer lo que el programa logró, cómo operó, y cómo fue juzgado por las personas involucradas y los expertos en el área del programa. Las preguntas más específicas de la evaluación surgen a medida que el estudio se desarrolla, basado en la continua interacción del evaluador con las distintas partes interesadas.

Stake ha expresado escepticismo en la investigación científica como fiable para desarrollar generalizaciones acerca de los servicios humanos, y pesimismo acerca de los beneficios potenciales de las evaluaciones formales de programas. Por lo tanto, la metodología y presentación utilizadas son básicamente cualitativas: métodos pluralistas, flexibles, interactivos, holísticos, subjetivos y orientados hacia el servicio. Dependiendo del propósito de la evaluación, el evaluador puede legítimamente emplear un rango de diferentes métodos, por ejemplo el estudio de caso, observación, reportes de los oponentes, y reportes narrativos. La presentación del informe suele ser informal y narrativa. La evaluación del programa puede culminar en hallazgos y conclusiones conflictivas, dejando la interpretación a los ojos del espectador.

La mayor fortaleza de este enfoque es que involucra la acción-investigación, en la cual la gente que financia, implementa y usa los programas, son ayudados a conducir sus propias evaluaciones y usar los hallazgos para mejorar sus conocimientos, decisiones, y acciones. Al mismo tiempo que los evaluadores analizan los intereses de los distintos actores

y buscan información relevante, también examinan los fundamentos, antecedentes/ trayectoria, procesos y resultados del programa (tanto intencionados como no intencionados). Su debilidad tiene que ver con la credibilidad externa, a raíz de que la gente en el contexto local tiene un control considerable sobre la evaluación de su trabajo. De igual manera, el trabajo tan cercano del evaluador con las partes interesadas puede perder la perspectiva de independencia. El enfoque no es muy dado a reportar los hallazgos a tiempo para la decisión o en la fecha límite para la rendición de cuentas. Además, más que conllevar la finalización, los aspectos contrarios y cualidades divergentes del programa pueden generar confusión y pugnas entre las partes interesadas.

6.8. Evaluación democrática deliberativa

Este enfoque planteado por Ernest House funciona dentro de un marco democrático explícito. Se propone emplear la participación democrática en el proceso de llegar a una valoración defendible del programa. El evaluador determina las preguntas de la evaluación, pero lo hace a través del diálogo y la deliberación con las partes interesadas. House y Howe presentan 10 preguntas para valorar la idoneidad de la evaluación democrática deliberativa: ¿Los intereses de quién están representados? ¿La mayoría de las partes interesadas están representados? ¿Alguien está excluido? ¿Hay serios desequilibrios de poder? ¿Hay procedimientos para controlar los desequilibrios? ¿Cómo participa la gente en la evaluación? ¿Qué tan real es su participación? ¿Cómo es su interacción? ¿Hay deliberación reflexiva? ¿Qué tan contemplada y extendida está la deliberación?

El enfoque considera esenciales, en todos los aspectos de validez de evaluación del programa, tres dimensiones: participación democrática, diálogo para examinar y validar las sugerencias de las partes interesadas, y deliberación para llegar a valoraciones defendibles del mérito y valor de un programa. La dimensión democrática, identifica y planifica una participación equitativa de todas las partes interesadas a lo largo del curso de la evaluación, no tolera un desequilibrio de poder. En la dimensión del diálogo, el evaluador encarga/involucra a las partes interesadas y a otras audiencias, a ayudar en la recopilación de hallazgos preliminares. Posteriormente, los colaboradores discuten y debaten los hallazgos del anteproyecto para asegurarse que ningún punto de vista de los participantes

está mal representado. En el escenario deliberativo, el evaluador considera y discute con otros todas las sugerencias e información obtenida, y solamente después presenta lo que considera una valoración del mérito y valor de un programa. A todas las partes interesadas se les da voz en la evaluación, y aunque el evaluador reconoce sus puntos de vista en el reporte final, también puede expresar su desacuerdo con algunos de ellos. Los métodos empleados pueden incluir charlas con las partes interesadas, encuestas y debates. La incorporación, el diálogo y la deliberación son considerados relevantes en todos los escenarios de una evaluación (concepción, diseño, implementación, análisis síntesis, redacción, presentación y discusión/debate).

Entre las ventajas del enfoque es posible mencionar que es un intento directo para hacer evaluaciones justas, y que se esfuerza por la participación democrática de las partes interesadas en todos los escenarios de la evaluación. Busca la incorporación de sus puntos de vista y emplea el diálogo para examinar y validar sus sugerencias. Al rendir la valoración final, el evaluador asegura la finalización. Su debilidad reside, tal y como lo han admitido House y Howe, en que este enfoque es, hasta ahora, poco realista y a menudo no puede ser aplicado de manera completa.

6.9. Evaluación constructivista

Guba y Lincoln (1985, 1989) son pioneros en la aplicación del enfoque constructivista de la evaluación de programas. El enfoque está orientado al servicio, emplea una epistemología subjetivista y rechaza la existencia de cualquier realidad final. El principal propósito es determinar y dar sentido a la variedad de construcciones que existen o surgen entre las partes interesadas. La investigación está abierta a la continua comunicación y a la recolección, análisis y síntesis de construcciones futuras. Una construcción no es considerada más “verídica” que otras, pero algunas pueden ser juzgadas como más informadas y refinadas. Todas las conclusiones de la evaluación son vistas como inconclusas y se exalta la apertura y continuidad en la investigación para encontrar mejores respuestas. Todas las construcciones son también dependientes del contexto. A este respecto, el evaluador define los límites de lo que está siendo investigado. El enfoque usa a los participantes como instrumentos en la evaluación, tomando así ventaja de sus experiencias relevantes, conocimiento, y perspectivas

de valor. El enfoque insiste en que el evaluador debe apoyar a todos los participantes, especialmente a los discriminados, mediante el aumento de su conocimiento, de tal manera que sean estimulados, informados y apoyados para transformar su realidad.

Las preguntas en la evaluación son establecidas por los evaluadores y las partes interesadas, y pueden surgir durante el proceso de formulación y discusión de la base lógica del estudio o durante la planeación de la agenda de debates. Las preguntas pueden cubrir, o no, el amplio rango de cuestiones involucradas en la valoración del mérito o valor de algo, y nunca son consideradas fijas. El enfoque hace uso de los métodos cualitativos. La metodología constructivista es primero divergente y después convergente. A través del uso de la hermenéutica, el evaluador colecta y describe construcciones alternativas individuales en una pregunta o cuestión de la evaluación. Los canales de comunicación se mantienen abiertos durante toda la investigación, y se otorgan facilidades a los informantes para que hagan sus sugerencias, al mismo tiempo que son informados de todos los aspectos del estudio. Para lograr el mayor consenso posible entre las diferentes construcciones, el evaluador cambia a un proceso dialéctico. Así, el evaluador involucra a los informantes en un proceso de estudio y confrontación de las construcciones existentes, reflexionando las diferencias entre las construcciones, y transitando tan lejos como se pueda hacia un consenso.

Entre sus principales ventajas está buscar la implicación directa de todas las partes interesadas que podrían ser perjudicadas o ayudadas por la evaluación. Con esto, tiende a ser educativo para todos los participantes, independientemente de que se alcance o no un consenso. Al no prometer respuestas finales, se mueve dentro de un escenario divergente, en el cual se busca ampliamente el entendimiento y el juicio, a un escenario convergente en el cual se buscan/postulan algunas respuestas unificadas. Las principales desventajas se encuentran en que la necesidad de una amplia intervención y la constante interacción a través de los escenarios divergentes y convergentes, hace difícil producir a tiempo los reportes que las agencias financiadoras y tomadoras de decisiones demandan. Además, si el enfoque es bien trabajado, requiere de la atención y participación responsable de un amplio rango de actores con intereses. El enfoque parece ser utópico en cuanto a su amplitud, intereses populares y participación, ya que a menudo son difíciles de obtener y sostener a lo largo de la evaluación del programa,

y más aún la situación puede ser empeorada por la continua rotación de las partes interesadas.

6.10. Evaluación de estudio de caso

Pioneros en aplicar el estudio de caso a la evaluación de programas fueron: Campbell (1975), Guba y Lincoln (1985) y Stake (1995). Una evaluación basada en un estudio de caso está enfocada en una descripción profunda, análisis y síntesis de un programa u objeto en particular. El propósito es proporcionar a las partes interesadas y a sus audiencias una explicación fiable, minuciosa y bien documentada del programa. Los investigadores no controlan el programa de ningún modo, más bien, ellos ven lo que está ocurriendo o lo que ocurrió en el pasado. El estudio ve el programa en su contexto geográfico, organizacional, cultural e histórico, examinando de cerca sus operaciones internas y el uso de los insumos y procesos para producir resultados. Examina un amplio rango de resultados, tanto esperados como inesperados. Define y describe las intenciones del programa y a los actuales beneficiarios, de estos últimos examina sus necesidades y las necesidades establecidas en el programa. Por lo tanto, al delinear y clarificar un programa, más que guiar su desarrollo o valorar y juzgar su mérito o valor, el enfoque de estudio de caso se tipifica como un enfoque orientado a las preguntas y métodos, más que como un enfoque de mejora y rendición de cuentas.

Para establecer las preguntas en un estudio de caso el evaluador debe identificar e interactuar con las partes interesadas en el programa y determinar las preguntas que sean de su mayor interés. Durante el proceso, las partes interesadas estarán involucradas para planear el estudio e interpretar los hallazgos. Las audiencias incluyen a los cuerpos de supervisión del programa, administradores, staff, promotores financieros y beneficiarios del programa. Algunas de las preguntas planteadas son: ¿cuál es la idea del programa? ¿Cómo ha evolucionado con el paso del tiempo? ¿Cómo opera en realidad para producir los resultados? ¿Qué se ha producido? ¿Cuáles son las deficiencias y los efectos secundarios negativos? ¿Cuáles son los efectos secundarios positivos? ¿En qué forma y en qué grado valoran el programa las diferentes partes interesadas? ¿A qué nivel el programa atendió efectivamente las necesidades de los beneficiarios? ¿Cuáles son las razones más importantes para el éxito o fracaso de un programa? ¿Cuáles

son las cuestiones más importantes no resueltas del programa? ¿Cuánto ha costado? ¿Cuáles son los costos por beneficiario, por año, etcétera? ¿Qué partes del programa han sido exitosamente trasladadas a otros sitios? ¿Cómo se compara este programa con sus competidores críticos?

Para un estudio de caso eficaz se necesita recurrir a un amplio rango de métodos cualitativos y cuantitativos, se puede incluir el análisis de archivos, muestras, análisis de contenido de documentos del programa, observaciones independientes y participantes, entrevistas, análisis lógico de operaciones, grupos de opinión, cuestionarios, escalas de valoración, test, audiencias, foros, mantenimiento de la base de datos del programa, etcétera. La veracidad se establece mediante el empleo y triangulación de múltiples perspectivas, métodos, y fuentes de información. Los reportes pueden incorporar descripciones y anécdotas de las tendencias históricas claves; centrarse en incidentes críticos, fotografías, mapas, testimonios, recortes de noticias relevantes, modelos lógicos y cuadros; y resumir las principales conclusiones. Para esto se puede utilizar tanto medios audiovisuales como documentos impresos.

Las ventajas de este enfoque radican en que ve el programa de una manera holística y minuciosa, dentro de contextos pertinentes, describiendo las influencias contextuales y examinando el trabajo interno y cómo producen los resultados. Incluye procesos claros para analizar la información cualitativa. La evaluación puede ser hecha a la medida para centrarse en las preguntas más importantes de las audiencias y puede ser retrospectiva o en tiempo real. La principal limitación del estudio de caso es que debido a la preocupación con la información descriptiva, el evaluador puede no recolectar la suficiente información evaluativa que permita una valoración completa del mérito y valor de los programas. Al tratar de producir una descripción completa del programa, el evaluador puede no producir oportunamente la retroalimentación necesaria para ayudar al desarrollo del mismo.

6.11. Evaluación de resultados como valor añadido

Como lo señalamos en el recorrido histórico, el mayor interés por este enfoque surgió en los años noventa del siglo pasado. Stufflebeam señala que este enfoque desarrollado por Sanders y Horn (1994); Webster (1995); Webster, Mendro y Almaguer (1994); y Tymms (1995), pone especial interés en el uso de pruebas estandarizadas para evaluar los efectos de los

programas y las políticas, con el propósito de proveer dirección para su elaboración, rendición de cuentas para los ciudadanos, y retroalimentación para mejorar los programas y servicios; al tiempo que asegura la estandarización de los datos para la valoración y mejora a lo largo de todo el sistema. Comúnmente se pone énfasis en las pruebas anuales o en la sucesión de los niveles escolares para valorar las tendencias y efectos parciales de los diferentes componentes de un sistema de educación (incluyendo grupos escolares o docentes). Al interesar más el conjunto que la actuación individual de los estudiantes, este proceso se caracteriza por la recolección anual de resultados medibles basados en indicadores estandarizados, por el análisis para determinar qué valor está siendo agregado al logro de los estudiantes, y por el reporte de los resultados para la rendición de cuentas, la mejora y los demás propósitos de la política. Así, un departamento estatal de educación puede recolectar anualmente datos del logro de sus estudiantes y el evaluador puede analizarlos para contrastar los resultados de las diferentes escuelas. Los resultados pueden ser analizados más allá para hacer comparaciones entre las áreas curriculares, maestros, escuelas básicas frente a escuelas medias, clasificaciones de la escuela, distritos y áreas del estado por tamaño y recursos.

Las preguntas establecidas por la evaluación se originan en las agencias de gobierno, organismos de financiación, responsables de tomar las decisiones, profesionales del sistema, y en los ciudadanos. Algunas de las preguntas que se hacen son: ¿a qué magnitud los programas agregan valor a los logros de los estudiantes? ¿Cuáles son las tendencias en los resultados a través de los años? ¿En qué sectores del sistema está trabajando peor o mejor el programa? ¿Cuáles son las deficiencias dominantes en los objetivos de un programa que requieren más atención y estudio?

La principal ventaja del enfoque es la sistematización y la institucionalización de la base de datos de los resultados (que puede ser usada con el paso del tiempo) y la manera estandarizada de estudiar y encontrar medios que mejoren los resultados. Así, hace asequible el análisis de tendencias a niveles del estado, distrito, escuela y salón de clases, haciendo uso eficiente de los test estandarizados. La mayor desventaja es que es políticamente cambiante al ser usado para identificar responsabilidades por el éxito y fracaso de los niveles escolares y maestros. Es altamente dependiente en información cuantitativa y, por tanto, el análisis complejo y eficaz está basado en un ámbito limitado de las variables de resultado.

Más allá de la emisión de medidas de resultado, el enfoque no provee una documentación exhaustiva de los insumos y procesos del programa y hace poco uso de los métodos cualitativos.

Como se puede observar, existe una amplia variedad de enfoques en el ámbito evaluativo. Los evaluadores y clientes pueden elegir, con base en las necesidades e intereses de la evaluación, aquellos enfoques que darán credibilidad a su trabajo. La mayoría de los enfoques presentados, muestra una fuerte orientación al hecho de involucrar a los *stakeholders* en el diseño y desarrollo de la evaluación; igualmente optan por el uso de múltiples métodos, es decir mezclar tanto métodos cuantitativos como cualitativos; así como que los resultados de las evaluaciones se utilicen e impacten en los programas o las organizaciones. Los enfoques evaluativos resaltaron igualmente los aspectos principales que tanto clientes como evaluadores, valoran más en cada enfoque, esto lleva a conocer lo que el cliente quiere y sus expectativas de la evaluación, y en este sentido la mayoría de los enfoques resaltaron la importancia de conocer el contexto en el cual se lleva a cabo el programa y la evaluación, pues ignorarlo puede llevar a un fracaso de la evaluación. Otro aspecto sobresaliente en la mayoría de los enfoques es la capacidad de negociación de los evaluadores para conciliar los intereses de los actores y mantener la integridad de la evaluación. En este mismo sentido, los enfoques resaltan el llegar a un acuerdo entre clientes y evaluador acerca de los principios y estándares que guiarán la evaluación.

Stufflebeam (2007) concluye que la debilidad de los estudios con orientación política es que son vulnerables a los conflictos de intereses y pueden engañar a una audiencia, al desarrollar tal vez juicios erróneos del mérito o valor de un programa. El principal problema con los estudios orientados a las preguntas y métodos es que a menudo establecen preguntas que son demasiado estrechas para apoyar una valoración completa del mérito o valor de un programa. Los estudios orientados a la mejora y rendición de cuentas, con su concentración en el mérito y valor, emprenden una tarea ambiciosa. Los estudios de agenda social y apoyo son reconocidos por su búsqueda de equidad así como de excelencia en los programas estudiados; tienen el reto de hacer de la evaluación una empresa democrática y participativa, algo utópico quizás. Por último, los enfoques eclécticos son atractivos por centrarse en hallazgos útiles y en su habilidad de atraer a las principales partes interesadas para diseñar y aplicar los marcos conceptuales, criterios y procedimientos de otros enfoques relevantes.

7. Tipos de evaluación

Los enfoques evaluativos ofrecen metodologías para realizar evaluaciones, sin embargo, necesitamos también delimitar la evaluación y “mostrar las diferentes posibilidades con las cuales podemos aplicar la evaluación y las virtudes que se obtendrán con ellas” (Casanova, 2002), con esto nos referimos a los tipos de evaluación. Decidir qué tipo de evaluación realizar está condicionado por las preguntas clave a las cuales queremos dar respuesta en la evaluación, por los objetivos y por los criterios de valor a utilizar. Una tipología de evaluación comúnmente utilizada es clasificarla en cuatro grupos: de acuerdo con la función que cumple la evaluación, al momento en que se evalúa, a los agentes que la realizan, y a la fase de contenido del programa que se evalúa (Osuna, 2000; Bustelo, 2001; Casanova, 2002, 2004). Estas tipologías, en palabras de María Bustelo (2001: 109), “son básicas, ya que son aplicables simultáneamente, es decir, cada evaluación podría ser clasificada por todas y cada una de estas cuatro clasificaciones (...), pudiendo darse una evaluación sumativa, de procesos y resultados, mixta y ex post a un mismo programa”. En este sentido, presentamos un breve repaso a cada una de estas tipologías.

7.1. La evaluación según la función que cumple

Esta tipología fue propuesta inicialmente por Scriven (1967), quien diferenció entre evaluación formativa y evaluación sumativa, convirtiéndose esta distinción en una de las más utilizadas en el ámbito educativo y en los demás ámbitos en general. Se refiere a si se quiere utilizar la evaluación para ir (con) formando el programa, servicio o política evaluada –evaluación

formativa– o si se quiere utilizar la evaluación para hacer recapitulación de dicho programa, servicio o política y realizar una valoración global de él –evaluación sumativa– (Bustelo, 2010).

La evaluación formativa implica la obtención rigurosa de datos a lo largo del proceso para que en todo momento se posea el conocimiento apropiado de la situación evaluada, que permita tomar las decisiones necesarias de forma inmediata (Casanova, 2002). La información se valora sobre la marcha, para que en caso de ser necesarios ajustes al programa para obtener los resultados deseados, estos se realicen. El objetivo principal es la mejora. En esencia, el programa se va “modificando y reforzando sobre la marcha como consecuencia de la evaluación formativa” (Bustelo, 2001: 110).

Por su parte, en una evaluación sumativa las preguntas y criterios de evaluación se establecen en el momento en que se determina cuándo y dónde se recogerá la información mientras se lleva a cabo el programa, servicio o política. “La información se recogería en dichos momentos, pero se iría acumulando para hacer un análisis y valoración conjunta al finalizar el programa o al terminar un ciclo del mismo. El resultado sería una evaluación global del programa durante el ciclo o periodo determinado que permitiría mejoras y modificaciones para ediciones posteriores” (Bustelo, 2001: 110).

7.2. La evaluación según el momento en que se realiza

Se refiere al momento en que la evaluación es llevada a cabo, y podemos hablar de evaluación *ex-ante*, intermedia o *ex-post*. La evaluación *ex-ante* se realiza antes de que la política, servicio o programa se implemente, analiza generalmente la adecuación de las medidas a las necesidades o problemática identificada (necesidades bien diagnosticadas; actividades, estrategias y objetivos pertinentes) determinando así el posible grado de éxito del programa. La evaluación intermedia es la que se realiza durante la puesta en marcha del programa, servicio o política. La evaluación *ex-post*, como su nombre lo indica, se realiza una vez finalizado el programa, servicio o política, y generalmente es el momento más común en el que se realizan las evaluaciones en los países con poca tradición evaluativa.

7.3. La evaluación según los agentes que la realizan

Una discusión tradicional surge cuando nos preguntamos quién debería realizar la evaluación. Hay quienes sugieren que la institución que diseñó y está implementando el programa debería realizar la evaluación (evaluación interna). La ventaja de esta evaluación es que al ser el mismo personal quien la efectúa, las personas están muy empapadas de los conocimientos y acontecimientos en el programa o servicio, por lo que existen mayores probabilidades de realizar la evaluación con información y datos pertinentes y adecuados, con lo que se esperaría mayores probabilidades de uso de los resultados de la evaluación. Sin embargo, esta misma implicación por parte del propio personal hace que la evaluación, a simple vista, no parezca muy objetiva y, por tanto, no tenga mucha credibilidad social. Ante esto, otros sugieren que instituciones o agencias ajenas al programa la realicen, para que así haya mayores posibilidades de objetividad y credibilidad social (evaluación externa). La dificultad aquí surgiría en la obtención de información pertinente. No obstante, una alternativa más recomendable sería que tanto el personal responsable de la ejecución y gestión del programa como personal ajeno al mismo, realicen la evaluación (evaluación mixta), aportando ambas partes su conocimiento en cada una de las distintas etapas.

7.4. La evaluación de acuerdo con las fases del programa

Esta clasificación toma como punto de referencia las tres principales fases del programa. Considerando, por una parte, que el nivel de desarrollo del programa condiciona en cierta medida el tipo de evaluación a realizar, en la etapa de planeación del programa, la evaluación del diseño y conceptualización del programa resulta particularmente adecuada. Cuando el programa lleva poco tiempo funcionando, la evaluación de la implementación parece ser la más apropiada y, la evaluación de resultados se corresponde con los programas que ya llevan tiempo funcionando. Por otra parte, las necesidades del cliente condicionan también el tipo de evaluación así, por ejemplo, si se busca saber si el programa está funcionando como se planeó, evaluar la implementación es lo adecuado.

Para realizar una buena evaluación no solo debemos dar cuenta de si se han alcanzado unos objetivos u obtenido ciertos resultados al final del

programa, es importante también conocer si cada una de las fases ha sido eficaz y por ende ha llevado a un éxito en el programa, conociendo así las posibles causas de éxito o fracaso y pudiendo actuar sobre ellas. Sin embargo, como menciona Monnier (1995: 269), “la realización de una evaluación requiere grandes dosis de modestia, puesto que hay que resignarse a no captar más que una parte de la realidad y a entender solo determinados aspectos de unos procesos complejos”.

La evaluación del diseño y conceptualización se realiza mediante un análisis del programa en relación con el problema sobre el cual se quiere actuar, para esto es necesario obtener información acerca del problema a solucionar, la población objetivo y el diseño del programa. Lo que se pretende es conocer la coherencia que existe entre el programa y el problema, ya que en el análisis se pueden detectar errores en el programa. La principal utilidad de este tipo de evaluación es conocer y asegurar una correcta conexión entre la teoría que sustenta la política y los instrumentos de la intervención (Tamayo Sáez, 1997). A este respecto, Carol Weiss habla sobre las Teorías de Cambio del Programa, saber qué espera alcanzar el programa y cómo espera alcanzarlo, esta Teoría de Cambio del Programa es la combinación de la Teoría del Programa y la Teoría de la Implementación.

Muchas de las veces la implementación define el éxito o fracaso de una política o programa, su buen diseño no es condición suficiente para un éxito efectivo. Es común que un programa no alcance los resultados esperados debido a que su puesta en marcha no se hizo de acuerdo con lo diseñado inicialmente, o no se tomó en cuenta a todo el entramado de agentes e instituciones que forman parte de la implementación del programa. Por consiguiente, es importante llevar a cabo una **evaluación de la implementación o de procesos**. Esta evaluación consiste en una descripción resumida del programa, en una recolección de información acerca del mismo (aplicación, funcionamiento, etcétera), para finalmente hacer una comparación entre el programa como está diseñado en la teoría y como realmente funciona, deduciendo así si la implementación fue o no adecuada (Alvira, 2002). Es decir, se estudian los procedimientos, actividades y rutinas, en general, todo el proceso que se utiliza para la ejecución de la política, tanto en su diseño como en la implementación, buscando emitir un juicio sobre la capacidad de procesamiento que tiene el sistema.

Para Freeman (1980) un proceso de evaluación está generalmente centrado en dos preguntas a las cuales debe responder: ¿está siendo el programa dirigido a la población adecuada o en el área objetivo? y ¿fueron las diferentes prácticas y esfuerzos de intervención llevados a cabo como se especifican en el diseño del programa o derivados de los principios explicados en el mismo? Y cómo la actividad evaluativa “implica dos aspectos interrelacionados, la evaluación de la política específica y los programas que la constituyen, y la evaluación de las personas que trabajan en las organizaciones responsables de la implementación de la política y los programas” (Parsons, 2007: 561), es en esta fase que se evalúa la gestión, lo cual supone un estudio de la institución que gestiona el programa, de sus procedimientos y criterios de funcionamiento interno (Osuna y Márquez, 2000).

La evaluación de los resultados consiste en conocer hasta qué punto los resultados fueron alcanzados y a qué costo (incluido tiempo, recursos humanos, materiales y financieros). Reconoce los efectos derivados de las políticas o programas aplicados, distinguiendo entre los productos (*outputs*) y los efectos (*outcomes*) de dichos productos, ya que muchas veces los resultados de un programa pueden ser achacados a otros factores externos o los efectos pueden repercutir en una población más grande de la población objetivo.

7.5. Otros tipos de evaluaciones

Estas tipologías expuestas no son las únicas que existen, dependiendo del grado de desarrollo del programa se pueden incorporar otros tipos de evaluaciones como la de necesidades, cobertura, impacto, económica, etcétera (Alvira, 2002). Sin embargo estas tipologías terminan por sobreponerse a las analizadas inicialmente. Aun así, hay algunas tipologías que vale la pena destacar por la importancia que cobran actualmente y que están relacionadas con los propios procesos evaluativos, nos referimos a la evaluabilidad y la Metaevaluación (Bustelo, 2001).

La idea de valorar la evaluabilidad de un programa, previo a su evaluación, fue propuesta por Wholey a mediados de los años setenta. La evaluabilidad, que debe estar establecida desde el diseño del programa, consiste en determinar la viabilidad del programa para ser evaluado. La finalidad es poder desarrollar una evaluación que responda a los propósitos

buscados, por lo que se busca conformar un programa que facilite una posterior evaluación. Existen tres características esenciales que determinan la evaluabilidad de un programa: la calidad de su planeación, el que exista la suficiente información tanto en cantidad como calidad que permita llevar a cabo la evaluación y, la capacidad de gestión de la Administración en relación con la evaluación. Si se cumplen estos aspectos se procede a diseñar y llevar a cabo la evaluación. La evaluabilidad es necesaria debido a que muchas veces los evaluadores se encuentran con resistencias políticas al momento de llevar a cabo una evaluación, o no; todos los programas pueden ser evaluados mediante metodologías cuantitativas. Así que para “evitar gastos y esfuerzos innecesarios” (Wholey, 1979) la evaluabilidad resulta indispensable antes de emprender una evaluación.

La metaevaluación es comúnmente definida como la evaluación de las evaluaciones. Scriven (1969) ¹⁵ acuñó este término cuando publicó un artículo en el cual respondía a las preguntas sobre cómo se evaluaban los instrumentos de la evaluación, usó la palabra metaevaluación para referirse a la necesidad de que las evaluaciones fueran a su vez evaluadas. Hoy en día, la metaevaluación parece ser una respuesta viable para mejorar la calidad de los procesos evaluativos, hacer un mejor uso de sus recursos y resultados, y promover una mayor comprensión de la tarea evaluativa. Stake (2006) comenta que del mismo modo que prestamos mayor o menor atención a las transacciones (procesos) del evaluando o a sus resultados (los productos), la metaevaluación puede poner un mayor énfasis en cómo se llevó a cabo la evaluación original o qué fue lo que produjo. La metaevaluación, de acuerdo con Vedung (1997: 41), puede ser llevada a cabo por el propio evaluador o por examinadores externos, con el objetivo de conocer la calidad, legibilidad, fidelidad a los hechos y otras propiedades de la evaluación. Refiere también que la metaevaluación puede ser un resumen sobre descubrimientos de diversos estudios de evaluación. A este respecto, María Bustelo (2001) distingue entre síntesis de resultados, meta-análisis y metaevaluación, donde esta última implica una evaluación de los procesos de evaluación, el énfasis está en los propios procesos de evaluación y no en los resultados arrojados por los mismos. Entonces, para María Bustelo, la metaevaluación significa la descripción, análisis y valoración de los

15 Para un repaso más amplio sobre los principales enfoques de metaevaluación citados en la literatura, consultar García Sánchez, E. (2009): Metaevaluación de políticas públicas: una visión desde la ciencia política. *Revista del CLAD. Reforma y democracia*, N. 43.

procesos de evaluación, además de la función de controlar de calidad de las evaluaciones ¹⁶.

Cabe resaltar también, la aportación que hace Ester García (2009) desde la ciencia política a la metaevaluación. Su propuesta de esquema se basa en dos premisas: 1. La metaevaluación ha de reconocer, analizar y enjuiciar tanto la dimensión técnica como la dimensión política de las evaluaciones; 2. Las decisiones que los distintos actores adoptan y las relaciones que se establecen entre ellos afectan e inciden en ambas dimensiones, configurando así los procesos de evaluación. La autora recalca que su esquema no pretende ser un catálogo de estándares que la evaluación debe cumplir, sino un conjunto de cuestiones o subdimensiones cuyo análisis permite reconstruir y enjuiciar el proceso evaluativo.

16 Para más detalle sobre la metaevaluación, ver María Bustelo “La evaluación de las Políticas Públicas de Igualdad de Género de los Gobiernos Central y Autonómicos en España: 1995-1999”.

8. El proceso de evaluación

La evaluación inicia con una serie de preguntas que nos ayudan a delimitar la evaluación (tipo), establecer la metodología (enfoque/s evaluativo/s) a utilizar y la función para la cual la realizamos. Para Stufflebeam y Shinkfield (2007: 473-474) una vez que se ha decidido realizar una evaluación, el evaluador necesita establecer un diseño apropiado, es decir, tomar las decisiones necesarias para llevar a cabo los requerimientos de la evaluación. Estas decisiones se centran especialmente en determinar el objeto a evaluar, las preguntas que serán establecidas, los criterios que serán aplicados, la información que será recogida, las herramientas de recolección de la información, análisis y procedimientos, el control de datos, los procedimientos de síntesis, reportes y métodos para los reportes, y pasos para promover y justificar el uso de los hallazgos. Este diseño de evaluación también debe establecer las necesidades de información de las audiencias, dar elementos para juzgar el mérito o valor del evaluando, realizar el modelo de evaluación elegido por el evaluador, ser capaz de ejecutarse en la realidad, y en general recoger los estándares de evaluación en el campo.

Para la evaluación, las políticas públicas y programas se presentan como objetos complejos que ofrecen distintos aspectos a ser evaluados, ya sea en forma separada o en conjunto, dependiendo de los intereses de quien solicita la evaluación. La evaluación se puede llevar a cabo a través de un esquema tradicional (Alvira, 2002) que no permite mucho contacto entre evaluador y evaluandos. Sin embargo, en la actualidad se han ido introduciendo esquemas más participativos, donde la participación comienza antes del diseño de la evaluación y ayuda a decir el tipo de evaluación a realizar,

entre muchas otras cosas. Lo primero es identificar a las partes interesadas y a los usuarios de la evaluación, después, ellos junto con el evaluador identifican las preguntas y los problemas así como los intereses y motivaciones que hay detrás de la evaluación (qué información se necesita) y en conjunto establecen los objetivos de la evaluación y las posibles vías de utilización de los hallazgos. La participación abarca también el análisis de la información y la redacción del informe de resultados, con el fin de hacer factible su utilización (Alvira, 2002; Osuna, 2000).

Sin importar lo que queramos evaluar, existen pasos en común a seguir al momento de llevar a cabo una evaluación. Es probable que estas etapas no se sigan al pie de la letra, pero su establecimiento nos sirve de guía para llevar a cabo una evaluación. Comentamos los trabajos de Fernández-Ballesteros (2001) y Stufflebeam (2007), ya que proponen un esquema del proceso evaluativo de seis fases y una lista de control para diseñar evaluaciones (*evaluation design checklist*), respectivamente. A continuación exponemos la lista de control para después señalar el proceso evaluativo, donde comentamos también aspectos de la lista de control en los apartados correspondientes.

8.1. Lista de control para diseñar evaluaciones de Daniel L. Stufflebeam (2007)

- A. Centrando la evaluación.
 - 1. Determinar y clarificar, el evaluando y el cliente.
 - 2. Identificar el mayor nivel de audiencias de la evaluación, por ejemplo, líderes del programa, personal y beneficiarios.
 - 3. Identificar de cada audiencia sus preguntas, sus necesidades de información y sus preocupaciones acerca de la evaluación.
 - 4. Identificar las partes que pueden ser perjudicadas por la evaluación y obtener sus *inputs*.
 - 5. Identificar los antecedentes de la evaluación y su contexto social y político.
 - 6. Identificar y establecer las barreras potenciales para la evaluación, por ejemplo, la necesidad de juntar información sensible, tener acceso a toda la información relevante, requerimientos de confidencialidad y anonimato, oponentes de la evaluación, conflicto de intereses, cuestiones de raza

y lenguaje, costos indirectos, y disponibilidad de los fondos necesitados.

7. Identificar y revisar información relevante, por ejemplo, evaluaciones previas del evaluando, evaluaciones de evaluandos similares, literatura pertinente y necesidades valorativas relevantes.
8. Acordar con el cliente sobre los estándares que guiarán y valorarán la evaluación.
9. Acordar con el cliente sobre el modelo o enfoque de evaluación que se aplicará.
10. Acordar con el cliente en el marco de tiempo, los evaluadores, las preguntas claves de evaluación, los reportes requeridos, las responsabilidades del cliente y las partes interesadas, y el costo permisible para la evaluación.
11. Aconsejar al cliente en invertir en una metaevaluación independiente.
12. Decidir si proceder con la tarea.

B. Recopilación de la información.

1. Considerar recolectar un amplio rango de información acerca del evaluando: contexto, historia, beneficiarios, benefactores, metas y estructura; contrastar con evaluandos similares, agendas, recursos, costos, personal, implementación, principales efectos, efectos secundarios, reputación, juicios de las partes interesadas y expertos, por ejemplo.
2. Escoger el marco para recolectar la información: estudio de caso, muestra de entrevista, experimento de campo, o un estudio multimétodo, por ejemplo.
3. Determinar las fuentes de información: documentos, expedientes, bases de datos, informes financieros, beneficiarios, personal, financiadores, expertos, oficiales de gobierno, o grupos interesados de la comunidad.
4. Determinar los instrumentos y métodos de recolección de información, por ejemplo entrevistas, observación participante, grupos focales, revisión de la literatura, búsqueda de archivos, Delphi, entrevistas, test de conocimientos, debates, visitas al sitio, fotografía, cintas de video, diarios, estudios libres de objetivos o estudios de caso.

5. Especificar los procedimientos muestrales para cada fuente: deliberado, probabilidad, o conveniencia.
6. Buscar establecer cada pregunta principal con múltiples métodos y puntos de datos.
7. Programar la recolección de información, indicar los tiempos cuando cada fuente de información y cada método será ocupado.
8. Asignar responsabilidades para la recolección de la información.
9. Dar al cliente y otras partes interesadas, una razón fundamental para el plan de recolección de información.
10. Revisar la viabilidad del plan de recolección de información con el cliente, y considerar la realización de reducciones prudentes.

C. Organización de la información.

1. Desarrollar planes y tareas para codificar, verificar, presentar, controlar y recuperar la información.
2. Diseñar una base de datos para la información obtenida, incluyendo el software apropiado.
3. Especificar equipo, instalaciones, materiales y personal requerido para procesar y controlar la información de la evaluación.

D. Análisis de la información.

1. Identificar las bases para interpretar los hallazgos, tales como las necesidades de los beneficiarios, objetivos, estándares, normas, los costos y actuaciones previas del evaluando, los costos y actuaciones de evaluandos similares, y juicios de expertos y de las partes interesadas en el programa.
2. Especificar los procedimientos de análisis cualitativo, por ejemplo, análisis temático, análisis de contenido, resúmenes, escenarios, o contrastación de fotografías.
3. Especificar los procedimientos de análisis cuantitativo, por ejemplo, estadísticas descriptivas, análisis de tendencias, análisis de costos; importancia de los tests para los efectos principales, interacciones, y efectos simples; meta-análisis;

análisis de factor; análisis de regresión, así como cuadros, tablas, y gráficas.

4. Seleccionar los programas de computación adecuados que faciliten el análisis cuantitativo y cualitativo.
5. Plan para buscar tendencias, patrones y temas en la información cualitativa.
6. Plan para contrastar diferentes subconjuntos de información cualitativa y cuantitativa para identificar hallazgos corroborativos o contradictorios.
7. Plan para establecer cada pregunta evaluativa por referencia y citando la información cualitativa y cuantitativa relevante.
8. Plan para usar la información cualitativa para elaborar y explicar los hallazgos cuantitativos.
9. Plan para exponer salvedades según proceda en consideración de todos los hallazgos no concluyentes o contradictorios.
10. Plan para sintetizar la información cuantitativa y cualitativa, por ejemplo, mediante incorporación de información cuantitativa dentro de una narración cualitativa o meter respuestas de entrevistas y otros hallazgos cualitativos en la discusión de hallazgos cuantitativos.
11. Anticipar que el cliente u otras partes interesadas pueden requerir recomendaciones para corregir los problemas identificados en los hallazgos, y estar preparados para explicar que los mismos datos que descubrieron el problema son improbables de dar dirección válida para resolver los problemas.
12. Considerar planear un proyecto de seguimiento que genere y valide las valoraciones de los cursos de acción alternativos para resolver los problemas identificados; tales procedimientos pueden incluir una evaluación de entrada de las estrategias alternativas de solución disponibles, creación y evaluación de nuevas estrategias de solución, compromiso de los principales expertos, revisión de la literatura relevante, o una conferencia de trabajo para trazar y valorar posibles cursos de acción.

E. Reporte de la información.

1. Aclarar a las audiencias los reportes de evaluación, por ejemplo, al cliente del programa, personal, la junta directiva de la política y los beneficiarios.
2. Identificar las necesidades del reporte por las diferentes audiencias, tales como provisionales, finales, o reportes con componentes específicos; reportes del contexto, entrada, procesos, y productos de la evaluación; apéndices técnicos; resúmenes ejecutivos; y un reporte interno de metaevaluación.
3. Para cada reporte, determinar los formatos apropiados, tales como impreso, oral, electrónico, multimedia, o narración.
4. Trazar las líneas generales del contenido de al menos los reportes principales, mostrando cómo los hallazgos de diferentes fuentes y métodos serán sintetizados para responder a las principales preguntas de la evaluación.
5. Considerar dividir el reporte final en tres subreportes: Antecedentes del programa (para aquellos que necesitan información sobre el pasado del programa); Implementación del programa (para aquellos quienes repetirán el programa), y Resultados del programa (para toda la audiencia).
6. En los apéndices técnicos, planear incluir resúmenes de la evaluación del personal y especialistas, información sobre los instrumentos de recolección y protocolos, reporte de hallazgos por un procedimiento particular de recolección de datos, tabla de datos, una lista de reportes provisionales, el contrato de evaluación, un resumen de los costos de la evaluación, y una justificación de qué tan bien el evaluador encontró los estándares de la profesión de la evaluación.
7. Desarrollar un plan y horario para entregar los reportes a las audiencias que tengan derecho a saberlo.
8. En su caso, obtener reportes preliminares de los proyectos de informe.
9. Conducir talleres de retroalimentación para ayudar al grupo de clientes en la revisión y discusión del boceto del reporte de evaluación.

- F. Administración de la evaluación.
1. Delinear calendario/programa de evaluación.
 2. Definir y planear las reuniones con el personal, así como con los recursos requeridos.
 3. Asegurarse que el plan de evaluación es suficiente para cumplir los estándares pertinentes de la evaluación de campo.
 4. Tener en cuenta al menos metaevaluaciones internas formativas y sumativas.
 5. Delinear un presupuesto para la evaluación.
 6. Negociar un contrato de evaluación, especificar las audiencias, responsabilidades del evaluador y protocolos, y redacción y difusión de responsabilidades y autoridad.
 7. Estipular revisiones y actualizaciones del plan de la evaluación y contrato en caso de ser necesario.

8.2. Planteamiento de la evaluación. Evaluabilidad del programa

La evaluación comienza con la petición de la misma y un estudio de viabilidad. Se realiza una primera recogida de información sobre el programa y sobre quienes dirigen la evaluación. La primera tarea del evaluador es situar al cliente (quien solicita la evaluación) en el contexto de la organización y la relación de esta con el programa para después clarificar, mediante entrevistas con los clientes, los propósitos que les han llevado a solicitar la evaluación. Especificar el objeto a evaluar resulta imprescindible, para lo cual es necesario analizar documentos relativos al programa y realizar entrevistas con responsables, gestores o técnicos, o visitar los lugares donde se implanta o ha implantado el programa. Con toda esta información, el evaluador está en condiciones de detectar los posibles obstáculos que pueden surgir, tanto procedentes de los implicados en la evaluación y el programa a evaluar como de los costos de evaluación, y saber si es posible llevar a cabo la evaluación (evaluabilidad del programa).

Antes de aceptar realizar una evaluación Stufflebeam (2007) también considera importante emprender una cuidadosa investigación preliminar para asegurar que será oportuno actuar. Esta investigación previa con las partes interesadas en la evaluación puede revelar al evaluador que no es posible llevar a cabo una evaluación responsable debido, por ejemplo, a

motivos ocultos del cliente. El evaluador debe desechar esas razones ilícitas para la evaluación y proponer opciones, o si lo considera oportuno rechazar la realización de la evaluación. Los 12 puntos del inciso A de la lista de verificación dan una guía para poner una evaluación justificable en terreno sólido. Para Fernández Ballesteros la evaluabilidad implica dos cuestiones: la calidad del programa (de su diseño, planificación, implantación y ejecución) y las barreras que pueden surgir a lo largo de la evaluación.

8.3 Establecimiento de los criterios de valor

El proceso de evaluación implica también el establecimiento de los criterios de valor. El emitir juicios de valor es la característica principal de la evaluación, si no se realiza esa acción estamos ante otra actividad como auditoría o investigación, por ejemplo, pero no evaluación. El control legislativo con las auditorías y el control de legalidad tratan de verificar si los recursos se han utilizado correctamente y de acuerdo a la normatividad, pero no valoran la eficacia ni los efectos de la actuación pública. Determinar el grado de eficacia de la actuación pública conlleva el establecimiento de juicios de valor previos, que no son los mismos para todos los implicados. La elaboración de los juicios de valor sobre el alcance de la acción pública, compete, de acuerdo con Monnier (1995), a una “instancia evaluadora pluralista” que asocia a los tres tipos de actores determinantes: quienes deciden, quienes ejecutan y los destinatarios.

Para poder realizar la valoración es necesario establecer criterios contra los cuales juzgar el objeto evaluado. Los criterios nos indican a qué características hemos de prestar atención (Stake, 2006: 116), mismas que se vienen perfilando desde el momento en que nos preguntamos sobre qué evaluar. Scriven sugiere desarrollar criterios de mérito que precisen qué tiene que hacer o cómo tiene que funcionar un programa para ser señalado como bueno, para cada criterio habrá que establecer estándares/normas de funcionamiento que especifiquen el grado de mérito. La evaluación será recopilar la información sobre el cumplimiento de estos criterios para valorar si se han alcanzado o no los estándares de funcionamiento.

Vedung (1997: 309) establece que determinar la valía de la intervención pública objeto de la valoración, es un proceso clave de la evaluación, y fija cuatro pasos para saber lo que constituye valiosa una intervención pública y cómo puede valorarse. Lo primero es identificar los criterios apropiados de

mérito a utilizar en la valoración. Sobre los criterios elegidos, seleccionar normas o estándares de rendimiento que denoten éxito o fracaso; después determinar el rendimiento real del evaluando sobre cada criterio y compararlo con cada estándar y, por último, integrar los juicios en una valoración única y global de la valía de la intervención.

Ha estado sobre la mesa de debate cómo establecer los juicios, se tiende a tratarlos como técnicas de medición para evitar involucrar valores e implicaciones políticas, pero la evaluación al realizar juicios de valor no puede evitar el componente político (como ya se ha visto en páginas anteriores). Durante mucho tiempo no se consideró el componente de valor de la evaluación debido a que parecía evidente que el evaluador debía evitar fijar sus propios criterios, siendo los objetivos enunciados en el programa los criterios naturales de mérito a aplicar. Parecía obvio que si los programas se ponían en marcha para conseguir determinados fines era lógico evaluar los mismos. Incluso la cuestión de unas normas o estándares podía resolverse de forma similar: si se alcanzaban los fines, los programas habrán tenido éxito y si no habrán fracasado (Shadish, Cook y Levinton, 1991: 46). Pero la realidad puso en evidencia que muchos programas carecían de objetivos específicos o que estos no eran operativos, que había objetivos no expresados en el programa y que un programa podía producir efectos no esperados, tanto negativos como positivos. Ante esto, otros criterios de mérito/valor fueron tomados en cuenta, por ejemplo, la satisfacción de las necesidades de usuarios; los efectos del programa tal y como son vistos por el personal técnico, administradores o usuarios; las necesidades de información de los interesados, las consecuencias del programa (Alvira, 2002). Estos criterios se relacionan más con enfoques pluralistas de la evaluación que con enfoques tradicionales.

De acuerdo con Tamayo Sáez (2002), los valores son el material para poder realizar juicios de valor, son los criterios para hacer la evaluación de políticas públicas y por tanto anteceden al juicio de valor, de modo que el evaluador tiene que adherirse a un ideal (un conjunto de valores) con el cual comparar las cosas o los acontecimientos. En el evaluador de políticas esos ideales pueden provenir de los valores pretendidos por los tomadores de decisiones, de los valores de los ciudadanos y grupos que mantienen relaciones de mayor o menor intensidad con las políticas públicas o de los valores que la sociedad y la comunidad política adscriben de forma consensuada al sistema político. Un estándar es una idea de lo que queremos,

son indicadores, referencias, puntos de comparación, criterios de medición para la verificación y contrastación entre una cosa planeada o esperada y una cosa obtenida o real y su carácter puede ser, por separado o a la vez, jurídico, administrativo y técnico (Mejía Lira, 2003). Garde (2005: 122) establece que para la selección de los estándares puede acudirse a políticas o servicios que destaquen por su excelencia, a las prácticas habituales en un determinado campo, a la referencia de expertos en la materia, a marcos legales, o bien pueden seleccionarse *ad hoc*, pero en cualquier caso los estándares obtienen su legitimidad en función de las tradiciones existentes, de las referencias normativas aplicables en la materia respectiva y del consenso que lleguen a alcanzarse entre los actores relevantes al efecto. Finalmente son el evaluador y el cliente quienes determinan cuáles serán más apropiados para la evaluación (Stufflebeam, 2007: 500).

De acuerdo con Shadish, Cook y Levinton (1991: 46), en la valoración descriptiva, el evaluador toma los valores de los demás como criterios y estándares. El planteamiento no es que esos valores sean lo más importante, sino son percepciones acerca de la validez de una intervención, materia prima para el proceso de adopción de decisiones. En la valoración prescriptiva, el propio evaluador “defiende la primacía de determinados valores”, por ejemplo, la justicia, la equidad y las necesidades del cliente, al margen de si esos valores han sido adoptados o no por un organismo encargado de la toma de decisiones o defendidos por algún grupo de participantes. Las teorías prescriptivas de valoración afirman que algunos valores son superiores a otros, mientras que las teorías descriptivas se limitan a describir los valores defendidos por los demás, sin refutarlos ni afirmar que un valor es el mejor de todos o mejor que algunos otros valores alternativos.

8.4. Selección de las operaciones¹⁷ a observar

Se trata de especificar o seleccionar las variables dependientes del programa y determinar el instrumento, procedimiento, técnica, test, medida o indicador con el que se dará cuenta de esas variables. Los resultados de las entrevistas, cuestionarios, tests, observaciones, datos de archivo, etcétera.

17 Las operaciones suelen ser la expresión de los objetivos que se pretende conseguir con el programa. Así, por ejemplo, en un programa educativo pueden referirse a las ejecuciones de una serie de test de rendimiento académico (Fernández Ballesteros, 2001: 29).

son operaciones observables en programas sociales, educativos y de salud. Las operaciones incluyen el *qué* vamos a observar o evaluar, *con qué* es observado y *desde qué* fuente de información. Para valorar los programas es recomendable utilizar distintos indicadores o medidas de múltiples variables seleccionadas y distintas fuentes de información, con el fin de lograr una correcta triangulación de la información. Stufflebeam aconseja al evaluador recopilar un amplio rango de información, como antecedentes del evaluando, estructura, actividades, costos, otros recursos y resultados, y también pide detalles con respecto de las fuentes de información, métodos e instrumentos.

Algunas de las cuestiones que resultan relevantes en esta fase del proceso, se refieren al establecimiento de las operaciones en las unidades¹⁸ que expresen mejor los cambios producidos en ellas como consecuencia del programa, dada la finalidad de la evaluación, los objetivos y las metas del programa. Al ser común que las operaciones sean la expresión de los objetivos del programa, estos siempre son un referente en la evaluación. Si el evaluador encuentra objetivos bien descritos en la planificación, significa que están implícitas determinadas operaciones a observar en distintos tipos de unidades, por lo tanto se pueden tomar esos objetivos bien especificados como operaciones a observar. Pero cuando los objetivos son muy generales, el evaluador tendrá que especificar, según unas medidas e indicadores concretos, los objetivos que se pretendieron conseguir. Puede darse también el caso de que los objetivos no hayan sido planteados ni de manera general ni específica y corresponde entonces al evaluador su especificación. Para eso puede utilizar como base entrevistas, cuestionarios o reuniones de grupo con los implicados, para después establecer los instrumentos de medida e indicadores de los objetivos establecidos por los distintos implicados.

Es en esta fase donde se plantea considerar otros efectos del programa potencialmente evaluables, ya sean positivos, negativos, deseables o indeseables. La bibliografía sobre el programa es útil para realizar las posibles predicciones. Las variables, instrumentos de medida e indicadores y fuentes de información utilizados para evaluar el programa en otros contextos, son de gran ayuda para conseguir un listado teórico de las variables dependientes u operaciones observables, así como para encontrar

18 Las unidades son el conjunto de sujetos a los cuales se ha implantado una determinada intervención, por ejemplo, individuos humanos, poblaciones, ciudades, escuelas, etcétera. (Fernández Ballesteros, 2001: 29).

los instrumentos y fuentes de información que pudieran ser los más indicados en la medición de dichas variables. Los distintos implicados en el programa y los expertos en el tema también pueden ser de gran ayuda.

Un aspecto a tener en cuenta es la accesibilidad a la información en los indicadores o medidas de los posibles cambios en el programa, puede ser que sean fácilmente accesibles o que haya que proceder a la construcción de instrumentos *ad hoc* para realizar la observación de las operaciones. En este último caso, la construcción de instrumentos depende del presupuesto disponible, cuando este es limitado el evaluador tendrá que analizar los resultados de instrumentos utilizados en estudios evaluativos previos, cuando el presupuesto lo permite el evaluador puede desarrollar procedimientos de medida altamente sofisticados, previa validación.

Debido a que las operaciones a observar se refieren a distintos implicados, es necesario establecer las fuentes de información. Debemos tener en cuenta que distintas audiencias merecen distinto grado de credibilidad, de modo que las entrevistas, observaciones, etcétera, se aplicarán a sujetos concretos. Los datos procedentes de archivos, documentos y pruebas no contaminadas, pueden ser tomados como indicadores y medidas de resultados de un programa.

8.5. Selección del diseño de evaluación

Un diseño es un plan desde el que se establece cuándo y qué unidades van a ser observadas (Fitz-Gibbon y Morris, 1987), su uso asegura que la recogida de datos sobre el programa se lleve a cabo de manera organizada. El diseño evaluativo es indispensable para contestar preguntas relevantes, fundamentalmente aquellas que exigen la comparación de distintas unidades sometidas, o no, al programa, o que les han sido administrados distintos programas que se desea comparar. Si durante la fase de planificación del programa no se ha establecido un pre-diseño de evaluación y no se ha recogido información para su posterior comparación, se tendrá que seleccionar el diseño para las operaciones a observar, el cual deberá también responder a las preguntas del cliente. Hay que tener en cuenta que los diseños alternativos dependen, entre otras cosas, de la finalidad de la evaluación, los medios de que se dispone, las características de las operaciones a observar seleccionadas y del tipo de unidad al cual ha sido aplicado el programa. Elegir un determinado diseño implica también la

selección de ciertas unidades, la cual estará en función de los objetivos de la evaluación y del programa.

Las principales amenazas a los diseños tienen que ver tanto con la validez interna como externa. La validez interna es que los resultados obtenidos mediante el diseño sean atribuidos al programa. Entre los factores a considerar destacan su historia, su consolidación, la administración de test, la instrumentación, la selección de las unidades, la mortalidad experimental o estadística y la interacción entre la maduración y cualquier otra fuente de sesgo (Campbell y Stanley, 1966). La validez externa implica que tales resultados puedan ser generalizados a otros universos. Los factores a considerar son: el efecto reactivo o de interacción de las pruebas, los efectos de interacción de los sesgos de selección y la variable experimental, los efectos reactivos del programa y las interferencias múltiples.

8.6. Recogida de información

Decididas las operaciones a observar y el diseño a utilizar, incluyendo las unidades a evaluar, se deciden los aspectos sobre la recogida de información. Éstos incluyen el establecimiento de la logística a seguir y de los sesgos potenciales que van a producirse en los datos a obtener. La logística se refiere a los pasos necesarios para una recogida de información adecuada (establecimiento de calendarios, formación de observadores, solicitud de autorización para la obtención de datos de archivo, visitas a informantes, etcétera) y dependerá de las decisiones ya adoptadas previamente respecto de los instrumentos/medidas/indicadores, o las potenciales fuentes de información (Fernández-Ballesteros, 2001: 97).

Stufflebeam, en los incisos B y C de su lista de chequeo, también pide escoger un marco apropiado de recopilación de la información, el cual a menudo supondrá una combinación de métodos. Pide establecer también los detalles de los procedimientos muestrales y cómo la información será combinada para responder a cada pregunta evaluativa. Hace referencia a la necesidad de desarrollar un programa para la recopilación de la información y asignación de responsabilidades para la recolección de la información, además toca el tema de justificar el plan de recolección de información y considerar si debería ser reducido. Estos puntos pueden ser considerados cuando se planea la evaluación, sin embargo las decisiones de muchos de ellos a menudo se hacen conforme va avanzando la evaluación. La

evaluación requiere también un enfoque eficaz de gestión de la información. En cada grupo de información se necesita seguir pasos sistemáticos para asegurar su precisión y garantía. Algunos datos serán codificados para su resumen y análisis posterior, y otros serán almacenados en computadora, y para todo eso el evaluador debe entrenar a quienes llevarán el trabajo, supervisarlos, regularmente checar la precisión en su trabajo, y asegurarse que solo personal autorizado accede y usa la información almacenada. El sistema de almacenamiento de la evaluación debe tener reglas claras y acuerdos para mantener la información segura y que garantice a los miembros autorizados del equipo de evaluación acceso a la información.

8.7. Análisis de datos: enjuiciamiento y recomendaciones

Esta fase consiste en sintetizar la información obtenida para responder a los propósitos de la evaluación. El tipo de análisis dependerá de la naturaleza de los datos y de las preguntas relevantes formuladas por el cliente de la evaluación o los propósitos de la misma. Brinkerhoff *et al.* (1983 en Fernández-Ballesteros, 2001) propone cuatro etapas relacionadas con los propósitos de la evaluación para realizar el análisis. La primera consiste en la revisión de las cuestiones planteadas, es decir, preguntarnos sobre la relevancia de las cuestiones formuladas a la luz de lo hallado hasta el momento, si están claramente formuladas como para guiar los análisis y si se pueden formular nuevas cuestiones. Lo siguiente es realizar análisis descriptivos, teniendo en cuenta que los análisis dependerán de la naturaleza de los datos (si los datos son cualitativos permitirán el análisis de contenido).

El examen de hallazgos y valoración sobre nuevos análisis, incorpora la revisión de los análisis realizados y la elaboración de nuevos supuestos y cuestiones con base en lo obtenido. Esto involucra la consulta a los miembros del equipo de evaluación y, en su caso, a implicados y expertos, con el fin de contrastar la relevancia de tales supuestos y cuestiones, para después preparar nuevos análisis descriptivos e inferenciales y así verificar los supuestos planteados, finalizando con la realización de los análisis programados. La última etapa consiste en la valoración de la evidencia final disponible, aquí resulta relevante preguntarse, por una parte, sobre la suficiencia de los resultados obtenidos para responder a las preguntas formuladas por el cliente o los implicados y, por otra parte, preguntarse si

dichos resultados sugieren la necesidad de profundizar mediante nuevos análisis y si es necesario volver a recoger nuevos datos y regresar a la etapa dos del proceso de evaluación.

Para Stufflebeam, el análisis en evaluación debe estar dirigido a responder las preguntas básicas de la evaluación y a juzgar el programa. Para desarrollar y sostener los juicios del objeto evaluado contra los criterios seleccionados, se requiere analizar de manera sistemática la información cuantitativa y cualitativa. Es necesario sintetizar los resultados del análisis de los diversos grupos de información, con el objetivo de combinar los hallazgos, a fin de responder a cada pregunta de evaluación y así emitir los juicios de valor. En las evaluaciones típicas, el resultado de la información y el análisis, no da bases sólidas para hacer recomendaciones sobre cómo mejorar el programa, porque el evaluador no recopila ni analiza información sobre el mérito, valor, y el costo efectividad de las recomendaciones contempladas. Bustelo (2001: 60) también añade que:

...la evaluación debe contar con este último nivel de análisis y exposición de resultados... las recomendaciones son sugerencias sobre posibles líneas de acción a seguir respecto al programa o política evaluada, cambios propuestos en el programa o cosas a mantener en el mismo o consejos para la mejora dirigidos a los patrocinadores, responsables y/o profesionales del programa. Estas recomendaciones deben seguir la secuencia lógica y sistemática de los niveles de análisis anteriores. Es decir, las recomendaciones de una evaluación deben estar basadas en las valoraciones realizadas que, a su vez, habrán estado basadas en la recopilación y análisis sistemático de la información.

8.8. Informe

Es la etapa final del proceso y consiste en la transmisión, ya sea oral o escrita, de los resultados obtenidos a las personas que solicitaron la evaluación, así como a todos aquellos interesados o potencialmente beneficiarios de la misma. Stufflebeam considera fundamental comunicar efectivamente los hallazgos a los interesados y asegurar el uso apropiado de la información reportada. Los pasos para promover y asegurar el uso de los hallazgos es una parte crítica en el diseño de la evaluación y necesita ser establecida

en la planeación de la evaluación, así como mantenerse a través de todo su proceso. Lo primero es identificar al cliente principal, pero también las preguntas y las necesidades de información pueden estar dirigidas a grupos más allá del cliente, por ejemplo, quienes tomarán las decisiones basados en los hallazgos, quienes tendrán la responsabilidad operativa de aplicar los hallazgos en la mejora del evaluando, y quienes están pagando por el uso de los servicios valorados.

Es vital estudiar todos los posibles públicos, con el fin de identificar sus diferentes necesidades de información y preparar la entrega de los reportes apropiados a cada uno de ellos. En general, los diferentes reportes pueden incluir reportes de componentes específicos tales como: de contexto, de entrada, de proceso, y de producto; reportes con un método particular, tales como encuestas, estudios de caso, o análisis de contenido; reportes del progreso intermedio; el reporte final, y apéndices técnicos. Dependiendo de las necesidades de cada público para cada reporte, el evaluador debe planear el empleo de una variedad de formatos más allá del reporte impreso. Para esto el evaluador debe interactuar con el cliente y demás grupos para conocer sus necesidades más importantes de información, motivarlos a recibir y usar los hallazgos, entregarlos en el momento adecuado, obtener su ayuda en la recolección de datos, y recibir las críticas y observaciones por los reportes previos. Stufflebeam resalta el cuidado del evaluador para mantener la independencia de la evaluación, así como recibir y valorar la retroalimentación de las partes interesadas (si es valiosa), pues no los debe empoderar para que tengan fuerte influencia en el diseño de las decisiones de la evaluación, ni consentirles ningún deseo o interés ilegítimo.

A menudo los reportes intermedios son tan influyentes, o incluso más, que el reporte final. Esto se da especialmente en las evaluaciones formativas, orientadas a la decisión y responsivas. Pero más allá de los reportes intermedios, los clientes requieren un reporte global final de la evaluación. Stufflebeam considera útil dividir el reporte en tres partes principales. Una sobre antecedentes del programa, útil a personas, quienes necesitan información sobre antecedentes del programa, incluyendo cuándo, por qué y cómo fue comenzado y por quién; su locación y ambiente, y la institución a la que pertenece. Una segunda sobre implementación del programa, de especial interés para grupos que pueden querer replicar el tema del programa. Esta parte debe ser altamente descriptiva más que evaluativa, debe identificar los objetivos del programa, beneficiarios, gobierno, personal,

organización, operaciones, y financiadores. El reporte de los resultados del programa debe ser dirigido a todo el público interesado y debe resumir la evaluación de diseño y proceso; presentar los hallazgos para cada pregunta principal, y sintetizar los hallazgos para presentar las conclusiones respecto del mérito, valor, significado, coste-efectividad, y probidad. Además de estas tres partes principales, el reporte debe incluir un conjunto de apéndices técnicos.

8.9. Administración de la evaluación

Es el último punto en la lista de chequeo de Stufflebeam, y se refiere a poder incluir en el plan inicial de la evaluación un plan de actividad de evaluación y de tareas del personal, que deben ser actualizadas apropiadamente durante la evaluación. El programa debe ser realizado junto con el cliente, para asegurar la disponibilidad de las partes interesadas en la recolección de los datos así como en su reporte, igualmente se debe incluir un adecuado presupuesto para las tareas de la evaluación. El contrato debe garantizar la viabilidad e integridad de la evaluación. El plan de evaluación debe considerar las revisiones y actualizaciones de los diseños de evaluación, y contrato en caso de ser necesario. Algo importante es que el evaluador debe reclutar, entrenar, y coordinar a los miembros del personal, de tal manera que lleven efectivamente todos los aspectos de la evaluación y se ganen la confianza del cliente y las demás partes interesadas.

9. La evaluación de diseño y gestión de programas

Tener claridad en el proceso a seguir al momento de desarrollar una evaluación, implica entre otras cosas, definir el tipo de evaluación a realizar y el establecimiento de los criterios con base en los cuales se emitirá la valoración de la política pública o el programa. En este apartado se señalan algunos criterios relevantes para evaluar el diseño y el sistema de gestión de los programas, por lo que a continuación se mencionan, previa exposición de la importancia de evaluar estas dos etapas en toda política o programa público.

9.1. La evaluación del diseño de programas. Concepto y objetivo

El que un programa cuente con un diseño adecuado es una de las condiciones para que se puedan dar los resultados esperados. El diseño no solamente implica los objetivos y la elaboración de pasos a seguir en el programa, sino también un estudio del problema a tratar, de sus antecedentes, además de un análisis de cómo se ha enmarcado el programa respecto de la legislación y de políticas, tanto nacionales como internacionales. Osuna y Márquez (2000) consideran que toda acción pública debería partir de un estudio riguroso de la realidad, es decir, del análisis de sus problemas y necesidades, y de sus causas y efectos, ya que a partir de su adecuada identificación es posible proponer objetivos y acciones que contribuyan a lograrlos.

Esto no puede ser de otra manera, los programas son aplicados en personas reales con problemas y demandas reales, y en contextos que cambian constantemente, así que no podemos diseñar un programa sin antes conocer la realidad y contexto en que vivimos. El objetivo principal de evaluar el diseño de un programa es analizar el programa en relación con el problema sobre el cual pretende actuar o, en palabras de Tamayo Sáez (1997: 308), “asegurar la correcta conexión entre la teoría que sustenta la política y los instrumentos de la intervención”.

9.2 Criterios para evaluar el diseño del programa

Esta fase de evaluación involucra el acercamiento al problema a tratar y el diseño del programa en sí mismo. Una de las metodologías para analizar esta etapa de la evaluación es la propuesta por Osuna y Márquez (2000), que parece clara y completa. Ellos proponen analizar tanto la racionalidad y la coherencia en el diseño del programa. Por racionalidad se refieren a la relación clara entre problemas, causas y objetivos, lo que significa, por una parte, evaluar el diagnóstico plasmado en el programa, es decir, su relevancia, y por otra parte, evaluar la calidad y la adecuación de la definición de los objetivos o su pertinencia. La coherencia hace referencia a la óptima correlación entre actuaciones, medios y objetivos, para ello es necesario evaluar la jerarquía en los objetivos del programa y la estrategia diseñada para su logro. El desarrollo de estas actividades comprende dos niveles de análisis: el primero se refiere a una coherencia interna, es decir, a la vinculación de los objetivos con los elementos de la estrategia y su concordancia con los problemas. El segundo nivel se refiere a una coherencia externa, es decir, a la coincidencia de los objetivos y la estrategia del programa con otras políticas y programas con los cuales pudiera ser complementario. Veamos de manera un poco más detallada y operativa estos elementos ¹⁹.

a. Relevancia.

Para evaluar la calidad del diagnóstico, que es de lo que se trata la relevancia, Osuna y Márquez (2000) consideran relevante el análisis de

¹⁹ Para un análisis más profundo, consultar Osuna, J.L. y Márquez, C. (2000): *Guía para la evaluación de políticas públicas*. Instituto de Desarrollo Regional, Fundación Universitaria, Sevilla.

las necesidades de la población objetivo y del contexto socioeconómico externo. El primer análisis incluye el tamaño de la población objetivo, su delimitación correcta, la densidad de los problemas que cubriría el programa y las potencialidades que pueden ser aprovechadas. Esta parte del análisis lleva implícita la correcta identificación de los problemas y fortalezas, mismas que deberán ser explícitas, claras, proporcionar una imagen verídica de la realidad de la población objetivo, definir de manera completa la realidad objeto de intervención, cuantificar las necesidades, y mencionar los documentos y personas que han sido consultadas para la elaboración del diagnóstico.

La delimitación de la población objetivo hace referencia al grupo de personas que cumple con las características para participar en el programa y su interés para colaborar en él. Este aspecto evalúa los problemas y potencialidades de la población objetivo, la cuantificación de la población, las características de la población para integrarse al programa y los criterios de selección, la cobertura del programa (congruencia en la definición de la población objetivo con las necesidades manifestadas, los recursos disponibles y los tiempos establecidos), y las barreras para participar en el programa (conocimiento y aceptación del programa y accesibilidad a las instituciones que brindan el servicio). Es deseable que existan documentos que contengan toda esa información, la evaluación se encargará de evaluar la calidad de la información contenida en ellos.

El análisis del contexto socioeconómico externo incluye todos aquellos elementos del entorno de la intervención (legales, políticos, económicos, institucionales, etcétera) que la afectan, ya sea de forma directa o indirecta. A pesar de que el personal involucrado en el diseño e implementación del programa, no puede controlarlos de manera directa, es necesario considerar estos factores para asegurar un mayor éxito en la intervención. Los aspectos que la evaluación valora son la consideración de análisis de contexto y de las variables que intervendrían en el problema, la realización de estudios de proyección que prevean la influencia del contexto externo en la implementación de programa y en sus resultados e impactos, y la rigurosidad y fiabilidad de las fuentes de información utilizadas.

b. Pertinencia.

La pertinencia se refiere al análisis de la calidad en la formulación de los objetivos. Los aspectos que se valoran tienen que ver con que los

verbos y enunciados usados en los objetivos expresen de manera clara las responsabilidades asumidas, para lo cual los objetivos deben ser medibles, observables, claros y concretos. Los objetivos deben manifestar una sola meta o fin (y por consecuencia un solo resultado o producto), la cual debe ser concreta y operativa. Finalmente, se valora que se hayan establecido fechas para el logro de los objetivos.

c. Coherencia interna.

Al hablar de coherencia interna, Osuna y Márquez se refieren al análisis de la estrategia. Uno de los primeros aspectos a analizar en este apartado es la jerarquía de objetivos. La lógica es que empleando ciertos insumos (económicos, técnicos, de personal, físicos, etcétera) se realizarán las actividades que permitirán alcanzar ciertos resultados, los cuales generarán cambios traducidos en el logro de los objetivos y fines del programa. Aquí se evalúa la hipótesis causal de intervención, sus variables deben ser medibles, de tal manera que muestren la influencia de las acciones sobre los aspectos que se busca modificar con el programa. Además de esta hipótesis, también se espera que haya una sobre los elementos del contexto externo que al momento de la implementación podrían afectar los resultados de la intervención. Se valora que en cada nivel de actuación no haya conflicto entre los objetivos, es decir, que exista complementariedad entre las estrategias de cada uno de los niveles. En este tenor, la evaluación valora que los objetivos específicos de cada nivel sean referencia para la intervención del nivel inmediato inferior.

El segundo aspecto de la coherencia interna tiene que ver con la relación entre problemas y objetivos, para esto debe haber una correspondencia entre los distintos niveles de intervención del programa, de tal suerte que no haya vacíos en la planificación y cada objetivo responda a un problema. Si esto es así, existiría la certeza de que cada uno de los objetivos del programa manifiesta una necesidad identificada en el diagnóstico previamente elaborado, por lo que el programa estaría bien diseñado en esta parte.

d. Coherencia externa en el diseño.

Este análisis valora la correspondencia que existe entre un programa y otras intervenciones que se implementan al mismo tiempo y en el mismo lugar. La evaluación implica la valoración comparada de los objetivos y las acciones de dichas intervenciones. Lo que pretende este análisis

es contrastar la complementariedad que existe entre el programa y las demás intervenciones y la posibilidad de una aplicación conjunta o, en su caso, la coordinación de las distintas intervenciones y su adecuación a la normatividad superior.

Estos elementos son propuestos por Osuna (2000) para evaluar el diseño de los programas. Sin embargo, Vedung (1997) en su propuesta de evaluación de proceso ²⁰ incluye factores que pueden influir en los resultados del programa, los cuales abarcan desde el proceso de formación de la intervención pasando por la aplicación de la misma. Vedung pretende dejar atrás aquella respuesta clásica de la tesis jerárquica donde las decisiones emanadas desde los altos niveles son perfectamente ejecutadas por los niveles inferiores, y según la cual no hay ninguna brecha entre las decisiones emanadas de la cúspide y los resultados reales. Esto último no sucede en la realidad puesto que cada decisión en cada etapa afecta a los resultados de una intervención pública. Vedung propone iniciar las evaluaciones de los programas con el análisis de los antecedentes históricos de la intervención para continuar con el diseño del programa. Se detalla a continuación los dos factores.

e. Antecedentes históricos de la intervención.

Para entender realmente cómo las intervenciones y su aplicación ejercen una influencia sobre los resultados, no basta con comenzar el análisis en el diseño propiamente dicho del programa, el análisis de los acontecimientos en la fase de formulación de políticas o programas, puede explicar por qué se ha obtenido determinados resultados. Existen seis subfactores en la historia legislativa de las intervenciones que tienen un peso especial: si la intervención institucionalizada está en línea o constituye una ruptura con medidas políticas previamente aplicadas; si la reforma se adoptó contando con un fuerte apoyo político, en situación de conflicto político o de consenso y cuál era la postura adoptada por los partidos y las corporaciones; si el cambio pretendía ser de grandes dimensiones o pequeño; si la intervención propuesta mereció mucha o poca atención durante su formulación; si la intervención se basaba en un propósito serio o si contenía rasgos de políti-

20 Para Vedung la evaluación de proceso tiene por objeto explicar con todo grado de detalle, porqué el programa ha condicionado los impactos o no; extrayendo conocimiento sobre la formación del programa, su aplicación y su marco general, como parte de un enfoque explicativo más amplio (Vedung, 1996).

ca simbólica, y finalmente, si esos organismos públicos, intermediarios privados y destinatarios centrales, regionales y locales que se verían afectados por el programa, participaron o no realmente en su conformación.

El primero de estos subfactores, la dirección del cambio propuesto, se refiere a que si determinada ideología política domina las ramas ejecutiva y legislativa del gobierno durante un prolongado periodo de tiempo, los funcionarios se acostumbrarán a trabajar para determinados tipos de personas, que se interesan por ciertos tipos de problemas. De eso se derivarán formas arraigadas de pensar y pautas o modelos inconscientes de actuación. Si entonces se produce un cambio ideológico entre los que están en el poder, por ejemplo, mediante un cambio del voto en unas elecciones públicas, los nuevos detentadores formales del poder descubrirán que no resulta fácil encauzar a los burócratas en la nueva dirección. Las dificultades de ajuste pueden llegar a convertirse en un obstáculo importante para la aplicación de nuevas medidas políticas.

El apoyo político, el segundo subfactor, se refiere a que los programas creados en condiciones de consenso partidista y corporativista son más fáciles de traducirse en resultados prácticos que los programas rodeados de conflictos. Son dos las situaciones conflictivas relevantes. En la primera, el programa es puesto en marcha por una mayoría contra la voluntad manifiesta de la oposición. En la segunda, el programa se transforma en una situación de compromiso entre los partidarios y oponentes originales, con lo que se asegura una mayoría. Si se adopta en condiciones de consenso político lo más probable es que el programa se aplique conforme a lo planeado, situación que no sucedería si se publicara después de conflictos políticos que no se han resuelto mediante la adopción de compromisos. El conflicto crea incertidumbre sobre el futuro del programa, lo que probablemente afectará a las instituciones en la fase de puesta en marcha. Incluso los destinatarios del programa pueden verse influenciados por el desacuerdo político.

La magnitud del cambio propuesto se refiere a que cuanto mayor sea el cambio propuesto, más difícil será la aplicación. Unas reformas pequeñas y graduales se llevan a cabo, probablemente, de manera más satisfactoria que grandes transformaciones revolucionarias. El nivel de atención prestado al programa por sus defensores en el proceso de formulación del programa, también puede influir en los resultados. Algunos participantes desaparecerán totalmente y aparecerían otros nuevos. Su nivel de atención

se verá afectado por la duración del periodo de adopción de decisiones, el número de participantes y la cantidad de cuestiones competitivas en el orden del día. Cuanto menor sea la atención prestada, más débiles serán la aplicación y los resultados. En cuanto a la participación en el proceso, si a los participantes afectados se les permite aportar ideas y argumentos para la formulación de la política, se dará origen a una cierta legitimidad que, por su parte, facilitará la aplicación y cumplimiento.

Una política simbólica significa que el programa se pone en marcha para fines distintos a los de conseguir resultados sustanciales. Los responsables de la adopción de las políticas puede que deseen ser vistos como firmes ante una situación, para satisfacer la opinión del partido o reforzar su propio liderazgo o reforzar la fidelidad de sus miembros, asegurarse votos en elecciones o facilitar coaliciones gubernamentales. El objetivo puede ser también minar la moral de los miembros de los partidos competidores, reducir su fuerza en elecciones, o debilitar su inclinación a formar coaliciones de gobierno. A efectos simbólicos, basta con iniciar programas y que no sean realmente llevados a cabo. Los funcionarios de los organismos afectados, los operadores de programa a nivel inferior y otros interlocutores del funcionariado pueden percibir que el programa tiene solo un contenido simbólico y dedicar menos energía a su aplicación de lo que habrían hecho en caso de no ser así (Saetren, 1983; March y Olsen, 1976). Barret y Fudge (1981: 276) al respecto comentan que:

La política puede convertirse en un sustituto para la acción, para demostrar que se está haciendo algo sin realmente abordar el auténtico problema... los gobiernos o los responsables de la adopción de medidas políticas desean que se les vea sensibilizados sobre una determinada cuestión, sin que ello implique necesariamente que quieran asumir la responsabilidad de la intervención.

f. Complejidad técnica y claridad del programa.

En esta etapa, Vedung se refiere a las características propias del programa y señala tres subfactores: uno es la complejidad técnica del programa, pues un programa altamente técnico resultará más difícil de entender y, en consecuencia, más difícil de aplicar que otro más sencillo. El segundo subfactor es la validez de la teoría del programa, de esta nos ocuparemos en el siguiente apartado.

El último subfactor se refiere a la claridad del programa. La ausencia de claridad se refleja en objetivos oscuros que dificultan a los implementadores y los destinatarios, formarse una idea correcta de que pretenden alcanzar las políticas. Esta oscuridad permite realizar cambios al programa en su concepción original. Muchas veces los documentos políticos contienen indicaciones indeterminadas y difusas, y toca a la evaluación clarificar qué se pretende con toda esa confusión. Hay oscuridad lingüística (palabras indeterminadas –ambigüedad y vaguedad–, y/o prioridades poco claras que ofrecen dos o más objetivos, pero sin indicación de a cuál de ellos se debería dar prioridad) y una oscuridad que puede ser lingüísticamente clara, pero consiste en diversas opciones para la acción (probablemente es el resultado de una solución y compromiso político para que el programa contenga, al menos, dos opciones de acción para los organismos encargados de su aplicación a alto o bajo nivel. La principal alternativa puede ser que los encargados de la aplicación deban actuar de determinada manera, pero que tampoco sea obligatorio hacerlo así, pudiendo actuar de manera contraria. La oscuridad política es con frecuencia intencionada y en un sistema democrático de gobierno va unida a juegos de poder político, o también puede surgir debido a limitaciones o escasez de conocimiento pertinente.

g. Teoría de cambio del programa.

Muchas veces los esfuerzos de un programa pueden no valer la pena si los supuestos en los que se basa son equivocados. Al igual que Vedung, Weiss (1998) pone el énfasis de la evaluación en no solo responder a la pregunta de si el programa ha funcionado o no, sino también busca conocer el camino que ha llevado a ciertos resultados, qué hizo funcionar el programa y a qué se debe que haya sido un éxito o un fracaso. Además, tampoco hay que limitarse solo a saber lo que el programa espera alcanzar sino también cómo espera alcanzarlo. La tarea del evaluador es conocer la realidad del programa y no tomar como válidos, a simple vista, las metas establecidas en el mismo, ya que muchas de las veces son confusas, vagas y ambiguas. Por lo tanto, Weiss sugiere que para realizar una evaluación que provea datos completos y reales, es necesario entender las premisas teóricas en las cuales está basado el programa, lo cual nos lleva a investigar la teoría implícita o explícita del programa.

Varios son los autores que han hablado acerca de la teoría del programa, entre ellos se puede mencionar a Rutman, 1980; Wholey, 1987; Monnier,

1995; Vedung, 1997 y Weiss, 1997 y 1998. Una intervención pública, al menos implícitamente, se basa en una concepción, por un lado, de los elementos determinantes del comportamiento problemático negativamente valorado a cambiar y, por el otro, la conexión entre el programa y estos determinantes (Pressman y Widavsky, 1984; Mazmanian y Sabatier, 1981, 1983; Sieber, 1981: 68 en Vedung, 1997: 238). Así entonces, para Vedung (1997), las teorías del programa constan de tres partes: la primera se refiere a la hipótesis causal sobre la influencia de uno o más elementos determinantes sobre el comportamiento que el programa se propone modificar; la segunda es la relación entre el programa y el determinante citado en la primera parte, como algo asociado con el comportamiento o condición a mejorar. La tercera es la parte normativa, es el valor positivo atribuido a reducir el comportamiento problemático.

Para empezar a definir la teoría del programa es necesario preguntarse qué ideas y suposiciones unen las entradas del programa al logro o consecución de los fines deseados. Por “teoría” Weiss se refiere al conjunto de creencias que están debajo de la acción, a las hipótesis mediante las cuales se construye el plan del programa. La Teoría del Programa sería el mecanismo que se presenta entre la entrega (y recepción) de un programa y sus resultados. La estructura operativa de cambio no corresponde a las actividades del programa por sí misma, sino la respuesta que esas actividades generan. Por lo tanto, la Teoría del Programa se enfoca en las respuestas de la población a las actividades del programa.

Pero el programa y su teoría implican también otra clase de suposiciones, lo cual tiene que ver con pasar del plano de las intenciones a las actividades del programa, es decir, se habla del plano de la implementación, la cual implícitamente también incorpora una teoría acerca de lo que se necesita para traducir los objetivos en servicios y operación del programa, a esta teoría Weiss la llama Teoría de la Implementación. El supuesto es que si las actividades son conducidas como se planearon, con suficiente calidad y fidelidad al plan, se lograrán los resultados deseados. La teoría de la implementación no se encarga del proceso que media entre los servicios del programa y el logro de las metas del programa sino se centra en la entrega de los servicios del programa.

Tanto la Teoría del Programa como la Teoría de la Implementación son complementarias. Cada escenario de actividad supone una reacción propia por parte de los participantes y la actuación de los participantes condiciona

el siguiente escenario del programa. Las dos corrientes de actividad tienen que interactuar para producir los resultados deseados. A esta combinación de las teorías es a lo que Weiss llama la Teoría de Cambio del Programa. Una teoría del programa usualmente incluye los insumos del programa, como recursos y apoyos organizacionales; las actividades del programa, las cuales representan la manera en la cual el programa es implementado, y los resultados provisionales, es decir, la cadena de respuestas que se obtiene de las actividades, las cuales se espera lleven a los resultados finales deseados.

La teoría debería trazar la secuencia de suposiciones que muestran cómo las entradas del programa (personal, recursos, actividades) se trasladan a una serie de pasos intermedios hacia los resultados deseados del programa (mejoramiento en las personas, organizaciones, o comunidades). Los micropasos de la teoría pueden ser el marco de trabajo para un estudio de evaluación. Los pasos de la evaluación se desarrollan con el propósito de conocer si los supuestos vínculos de hecho ocurren.

h. Marcos interpretativos.

Vedung considera importante antes de hacer un análisis del diseño del programa, elaborar un análisis de los antecedentes históricos del mismo para entender la intervención, su aplicación y su influencia sobre los resultados. La teoría del análisis de los marcos interpretativos sigue un poco este mismo camino, y aunque esta teoría se ha aplicado generalmente al ámbito de los movimientos sociales, Bustelo y Lombardo (2006) y Viñas (2009) lo han aplicado al ámbito de las políticas públicas, en los temas de género y a la catástrofe del Prestige, respectivamente. Así que la teoría de los marcos interpretativos puede ser aplicada también a diversos problemas públicos para conocer cómo se enmarca el problema a nivel internacional y nacional, y su repercusión en el diseño de programas.

El análisis de marcos interpretativos plantea como foco de análisis los documentos, nacionales e internacionales, que han dado forma a las políticas que se llevan a cabo en cierto país o estado, buscando la congruencia entre lo que existe como problema real, lo que los gobernantes perciben y lo que se hace para solucionar el problema. Un marco interpretativo de política (*policy frame*) es “un principio de organización que transforma la información fragmentaria o casual en un problema político estructurado y significativo, en el que se incluye, implícita o explícitamente, una solución” (Verloo, 2005: 20 en Bustelo y Lombardo, 2006).

En este sentido, Viñas (2009: 123) señala también que “el análisis de marcos ha demostrado ser un método de gran utilidad para estudiar el proceso de las políticas públicas, y el papel que en este proceso juegan los actores institucionales, políticos y sociales, ya que se centra en la construcción y reconstrucción de la realidad que realizan estos actores a través de sus documentos y discursos”. Agrega, que contribuye así a comprender la dinámica del proceso de formulación de esa política, además de demostrar cómo interpretaciones y perspectivas competitivas pueden llevar al diseño de políticas públicas diferentes. Para ella, enmarcar es una forma de seleccionar, organizar, interpretar y dar sentido a una realidad compleja para proveer puntos de referencia, a fin de conocer, analizar, persuadir y actuar. Pues en el análisis de políticas públicas no solo es relevante qué problemas han sido incluidos en la agenda pública, sino cómo se ha definido ese problema.

En este sentido, para Viñas la misma información puede originar definiciones y explicaciones conflictivas sobre un problema. En el contexto de las políticas públicas, un “problema” generalmente se estructura en un diagnóstico (cuál es el problema) y un pronóstico (cuál es la solución) de una cuestión determinada. En la construcción del problema pueden competir diferentes interpretaciones sobre el diagnóstico y el pronóstico. Dentro de las diversas definiciones del problema están, implícita o explícitamente, las representaciones referidas a quién tiene el problema, quién tiene la solución, cuáles son las causas del problema y cuáles son los medios para solucionarlo.

El grupo MAGEEQ²¹ desarrolló una metodología del análisis de marcos interpretativos. Se asume que “la existencia de <<marcos interpretativos>> (*policy frames*) diferentes en relación con el problema de la desigualdad de

21 MAGEEQ (Mainstreaming Gender Equality in Europe) es un proyecto de investigación, iniciado en 2003, dedicado a analizar y mejorar las políticas de igualdad de género de la Unión Europea y de sus países miembros, y concretamente la aplicación del concepto *mainstreaming* de género. Su método característico es el análisis de marcos interpretativos aplicados, en este caso, a textos políticos relacionados con cuatro temas seleccionados para el estudio (violencia de género, conciliación de la vida laboral y familiar, desigualdad de género en la política y derechos de homosexuales). MAGEEQ ha analizado los textos oficiales de las principales instituciones políticas en la Unión Europea (planes de igualdad, leyes, debates parlamentarios) y en los seis países miembros elegidos, así como textos producidos por el movimiento feminista y los medios de comunicación. El proyecto está financiado por la Comisión Europea en el ámbito del V Programa Marco.

género, influye en la formulación de las políticas” (Bustelo y Lombardo, 2006: 13). Así, diferentes significados atribuidos a determinada palabra, afectan a la manera en que se enmarca cierto problema y las soluciones para ello. Lo que se busca es distinguir las diferentes maneras en las cuales se enmarca un determinado tema. MAGEEQ asume que todo marco de política tiene un diagnóstico del problema y un pronóstico o solución, sin embargo, el grupo ha añadido tres elementos más a esta estructura básica de diagnóstico y pronóstico: voz, equilibrio entre diagnóstico y pronóstico (o distinguir si predomina uno u otro), y las normas (analizar lo que se considera explícita o implícitamente como bueno o malo en las posiciones normativas). Lo que propone el análisis de marcos interpretativos es:

...identificar la lógica interna (implícita o explícita) de los <<marcos interpretativos>>, poniendo énfasis en las diversas representaciones que los/as actores/as políticos/as ofrecen del problema de la desigualdad de género y las soluciones propuestas al mismo. (...) pretende provocar una reflexión acerca de la (in) coherencia de las políticas de igualdad y de las mejoras que podrían derivarse de un diseño más consistente de las mismas (Bustelo y Lombardo, 2006: 14).

Bustelo y Lombardo (2006: 18) buscan y analizan “las inconsistencias detectadas en la manera de enmarcar el problema y su solución. Destacan los desequilibrios existentes en los distintos temas entre el diagnóstico y el pronóstico del problema y las consecuencias que estos tienen para las políticas”. El análisis de tales documentos se realiza a través de una serie de preguntas aplicadas para todos los textos seleccionados, esto en una primera fase. En una segunda fase son estos análisis los que pasan a un segundo análisis, todos cualitativos, y del cual ya se obtiene el código o marco interpretativo dominante.

9.3 La evaluación del sistema de gestión e implementación de programas. Concepto y objetivo

El que un programa cuente con un buen diseño es una condición necesaria, pero no suficiente para el éxito y la obtención de los resultados esperados. Después de esta fase se pone en marcha otra que involucra la actuación del

personal encargado de gestionar el programa y de las personas a quienes va dirigida la acción. Nos referimos a la puesta en marcha de las acciones propuestas o implementación. El interés por la implementación surgió a raíz de la misma preocupación que la evaluación, debido al fracaso de los programas gubernamentales de los años setenta en los Estados Unidos, los cuales se consideraban perfectos, y quizá lo eran en el diseño, pero no tomaban en cuenta las dificultades que la puesta en marcha del mismo acarrearía. Se creía que una vez superada la etapa de decisión, así como de haber diseñado unas acciones “perfectas” y transmitir las a los administradores y administrados, estos las gestionarían o implementarían de la manera estipulada. Este supuesto derivó, además de costos excesivos, en que las acciones puestas en la práctica y los resultados obtenidos fueran muy diferentes de las políticas diseñadas y de los objetivos planteados.

Uno de los trabajos pioneros que abrió camino en la investigación de la implementación fue el realizado por Pressman y Wildavsky en 1973, bajo el título de “Implementación. Cómo las grandes esperanzas de Washington son destrozadas en Oakland, o por qué es sorprendente que los programas federales puedan llegar a funcionar aunque solo sea un poco”. Otro de los trabajos más citados es el de Bardach (1977) “El juego de la Implementación. ¿Qué sucede después de que un proyecto se convierte en ley?”. La importancia de estos estudios radica en el planteamiento del problema de la puesta en marcha de las políticas y programas públicos. Pusieron en evidencia la gran cantidad de redes y niveles de actores que participaban en la puesta en marcha de un programa, destacaron las relaciones y negociaciones que se daban entre ellos, y resaltaron la aceptación y disposición necesaria por parte de los involucrados en la implementación del programa para que este funcionara.

El análisis a varios programas sociales, puso en evidencia que los problemas de implementación se derivaban del supuesto por el cual los políticos y altos funcionarios toman la decisión, considerando que el curso de acción a seguir por la política era el que ellos diseñaban y que sus acciones técnicas, organizativas y operativas se desarrollarían al pie de la letra. Suponían que dichas acciones irían descendiendo desde los niveles centrales, pasando por los estatales y locales, hasta llegar, en términos de Lipsky (1977), a los “burócratas de nivel de calle” de manera inalterada. Este modelo conocido como *top-down* se basaba en la estructura jerárquica y en el concepto de burocracia de Max Weber, el cual coloca en la cima de

la pirámide a los tomadores de decisiones, seguido por los administradores que aplicarán esas acciones y en la base los destinatarios de las acciones del programa o política.

Como respuesta surgió un enfoque que puso la atención en los actores que prestaban el servicio al público. Finalmente, de ellos dependía el éxito o fracaso de todas las reformas propuestas por los gobiernos. Este enfoque fue conocido como *bottom-up* y centró su atención en el papel de los “burócratas a nivel de calle” en la implementación. Dentro de los burócratas a nivel de calle es posible incluir a los maestros, médicos, trabajadores sociales, policías, abogados, etcétera, es decir, “todos aquellos que en su trabajo interactúan directamente con los ciudadanos y gozan de gran discrecionalidad en el desempeño de sus funciones” (Lipsky, 1977: 172). La falta de claridad en los fines y objetivos, y los limitados recursos en comparación con las acciones que tienen que realizar, hacen que estos actores tengan que actuar con discrecionalidad al momento de prestar el servicio. “Para ello rutinizan procedimientos, modifican objetivos, racionalizan servicios, determinan prioridades y limitan o controlan clientelas” (Lipsky, 1977: 172).

El adentrarnos en evaluar esta fase del programa implica, de acuerdo con Subirats (1994: 140), preguntar:

... sobre el nivel de racionalidad de su organización interna, sobre el grado de predictibilidad de la actuación del personal implicado, sobre las pautas de dirección seguidas y la claridad en los niveles de responsabilidad, sobre el nivel de respuesta o de reacción que la aplicación de la política genera en la misma organización, o sobre la adecuación de los recursos destinados a esta política con los objetivos planteados.

Será importante conocer además “cómo se ha producido la implementación, es decir, cuál es el grado de operatividad del programa, las diferencias de aplicación que se hayan podido producir en distintos momentos o lugares, cuál ha sido la reacción de actores implicados, qué errores o abusos se han producido, etcétera. Es importante también conocer cómo ha funcionado la gestión del programa” (Subirats, 1994: 144-145).

El objetivo de evaluar la implementación de un programa es analizar hasta qué grado se han realizado las acciones estipuladas en el diseño

del mismo, conocer dónde están los aciertos o errores, y con base en esa información emitir un juicio y ayudar al proceso de puesta en marcha y gestión de las políticas o programas. Osuna y Márquez (2000) establecen que para que un programa logre los efectos esperados debe contar con un adecuado sistema de gestión y de seguimiento. La gestión de los programas implica todos los procesos relacionados con la puesta en marcha de las acciones de los programas y el logro de sus objetivos, así como los relacionados con el personal. El seguimiento será la verificación periódica y constante de las actividades realizadas, tanto física como financieramente.

El seguimiento o monitorización se considera más bien como una actividad administrativa permanente. El objetivo es producir información útil para la evaluación de la implementación, y también para evaluar los posteriores resultados e impactos. La monitorización se lleva a cabo de manera continua y permanente, e implica la definición de criterios operativos de logro y la construcción de indicadores –de procesos de gestión, de rendimiento, de cobertura, de uso de servicio y económicos– que se aplicarán sistemáticamente mientras se desarrolla la intervención (Tamayo Sáez, 1997), se realiza de forma interna. El seguimiento no implica un enjuiciamiento sistemático, a diferencia de la evaluación que puede realizarse de manera interna o externa, y cuyo objetivo es emitir un juicio sobre la validez de la intervención.

9.4 Criterios para evaluar la gestión del programa

Osuna y Márquez (2000) proponen una metodología para evaluar los sistemas de gestión en los programas, y plantean el análisis separado de los aspectos que condicionan la puesta en marcha de un programa. Ubican cuatro áreas a evaluar: la racionalidad de los objetivos y estrategias de gestión diseñados; la organización interna de la división de tareas y responsabilidades; el análisis del sistema administrativo para la difusión de las actividades del programa y para la recepción, y tramitación de expedientes; finalmente, la valoración de los procedimientos para la recogida de información sobre los avances y los retrasos en actividades realizadas. Para cada una de estas áreas establecen variables que condicionan su desarrollo óptimo. A continuación hacemos una revisión de ellos.

a) Racionalidad de los objetivos y estrategias de gestión diseñados.

El resultado de un programa dependerá del adecuado diseño del sistema de gestión y de su puesta en marcha. Por lo tanto, el órgano administrativo encargado de implementar el programa, necesita establecer unos objetivos y una planificación ordenada. Los aspectos a considerar en la evaluación son: claridad de los objetivos en cada nivel de gestión, y la adecuación de etapas administrativas y funciones establecidas, considerando si estas son o no las necesarias.

b) División de tareas y responsabilidades entre los gestores y ejecutores. Esta valoración incluye el análisis y la coordinación de diferentes dependencias administrativas involucradas en el seguimiento, la implementación y la evaluación de la intervención. Se pretende conocer si las funciones de cada una de estas dependencias están claramente diferenciadas, si existe duplicidad o ambigüedad en sus funciones, y si el personal conoce su función y aportación en todo el proceso.

Se valora la flexibilidad de los procedimientos organizativos para evaluar el logro de los objetivos y la coordinación en la toma de decisiones de los gestores. Los aspectos a analizar son las responsabilidades que cada área tiene en materia de implementación, seguimiento y evaluación, y la articulación de estas actividades en los organigramas. Las actividades y funciones de cada puesto relacionadas con la implementación, seguimiento y evaluación, deben ser claras.

Todo lo anterior involucra, por una parte, la valoración del personal y de todo aquello que tiene que ver con los sistemas de incentivo y castigo, con los servicios profesionales de carrera, con las actividades a realizar, etcétera. Es importante esta evaluación porque considera temas como la infrautilización del personal y su motivación en la gestión y la ejecución del programa, su cualificación para desempeñar las tareas que le corresponden y los métodos de trabajo que emplean. Por otra parte, implica la evaluación de recursos económicos, materiales y temporales empleados. Esto se valora, por ejemplo, con los medios que se destinan a las necesidades existentes, el tiempo que se dedica para realizar las tareas, y el presupuesto asignado para realizar las actividades de gestión y seguimiento.

Un factor para que la estrategia de cambio en las organizaciones se vea como una transformación en las actitudes y tareas de los gestores, y no con fines publicitarios o propagandísticos, es la voluntad política para

implementar un sistema de gestión. El interés en el proceso por parte de todo el personal es importante para que las funciones de gestión, seguimiento y evaluación, se conciban y conviertan en una nueva cultura administrativa.

c) Promoción y difusión de un programa.

El publicitar las actuaciones incrementa la información y conocimiento de la población a quien va dirigido el programa, esto fomenta su participación y la transparencia en la implementación de la intervención, y en consecuencia aumentan las posibilidades en el logro de los objetivos. Aquí se puede diferenciar dos niveles distintos de actuación: el primero estaría orientado a la publicidad del programa en general; el segundo, a las actividades más concretas de promoción y difusión de cada una de las actuaciones del programa (plan de comunicación).

Los aspectos que debe contener un plan de comunicación y que se deben analizar son: en primer lugar, el presupuesto destinado para la implementación de las distintas acciones del programa, ya que este determina el alcance de la intervención; por ejemplo, en el nivel de cobertura. Un aspecto particularmente interesante es el análisis entre líneas del reparto financiero, a través del cual se puede averiguar, por un lado, el peso que tiene cada acción, sobre todo el proceso de la intervención y, por otro, conocer la estrategia implícita.

Un segundo aspecto a evaluar son los criterios de selección y requisitos, que deben servir para que el personal valore los proyectos de acuerdo con los fines perseguidos en programa. El conocer de manera detallada los requisitos y criterios de selección resultará muy útil para que los beneficiarios potenciales del programa participen, o no, en él.

Un tercer aspecto es el horizonte temporal para la presentación de solicitudes, en la convocatoria deben quedar claramente establecidos el procedimiento, las fechas y el lugar donde se presentarán las solicitudes.

Esta etapa de la evaluación incluye también una parte más administrativa-burocrática en la cual se analiza la existencia y calidad de diversa información, como manuales o guías de procedimiento para la gestión del programa, los formatos de solicitudes de ayudas que incluyan la información necesaria para conocer el proyecto, el apoyo a los posibles beneficiarios para la elaboración de su proyecto, un proceso de recepción y registro de solicitudes, las bases de datos, los informes de evaluación de las solicitudes, y los sistemas de comunicación con los beneficiarios.

d) Mecanismos y procedimientos para la recogida de información.

Constituye la última área a evaluar dentro de la gestión del programa y se refiere a la cantidad y calidad de la información y a la forma en que se genera, Osuna y Márquez lo denominan seguimiento. Como se ha mencionado, esta constituye una actividad permanente, cuya información sirve para realizar la respectiva evaluación de la implementación. Su tratamiento requiere un apartado especial, que no es objeto de este libro.

Ahora bien, retomando los factores explicativos en la evaluación de proceso, Vedung (1997) toma como tercera y cuarta dimensión a estudiar, la aplicación del programa y la respuesta de los administrados hacia el mismo, respectivamente. Para la evaluación de esta tercera dimensión, Vedung se basa en la apreciación de Lennart Lundquist (1987: 76 ss.), para quien la aplicación implica comprensión, capacidad y predisposición por parte de todos los involucrados en el programa. Así entonces, Vedung propone que estos tres preceptos se apliquen al análisis de los organismos nacionales, los intermediarios formales, la burocracia a nivel de calle, y a la participación de los administrados.

e) Aplicación del programa.

Todas las personas involucradas en el programa deben ser capaces de comprenderlo, ser capaces de ejecutarlo y estar dispuestos a hacerlo. Una deficiente comprensión del contenido del programa, por parte de quienes tienen que aplicarlo, puede provocar problemas de ejecución. Por muy claro y coherente que haya sido formulado, un programa que no sea conocido por los solicitantes en potencia no podrá contribuir a alcanzar los objetivos para los cuales fue creado. El personal debe poseer las capacidades administrativas necesarias, es decir, poseer el conocimiento técnico, adaptarlo al programa y a las necesidades de los destinatarios. Los resultados también se ven influenciados por la predisposición administrativa, esto implica varios supuestos: el primero de ellos tiene que ver con que los directores o funcionarios persigan sus propios intereses, lo cual puede hacer retorcer el mandato político para adecuarlo a sus propios fines; otro supuesto es que los funcionarios responsables de la aplicación no aprueben los objetivos de la medida o le den baja prioridad respecto de los recursos, debido, principalmente, a cuestiones políticas; o que pongan en duda la efectividad de la acción y cuestionen las soluciones técnicas elegidas. Un último supuesto se refiere a que los destinatarios, en

ocasiones, no desean modificar ciertas acciones, sin embargo pueden optar por oponerse al programa o “captarlo”²².

Para lograr sus propósitos, los gobiernos recurren a intermediarios formales, quienes son agentes o partes no dependientes del gobierno (por ejemplo las ONG’s), que actúan como intermediarios o burócratas ocultos entre los organismos administrativos nacionales y los grupos destinatarios. El problema radica en que los intermediarios pueden apropiarse del programa y retorcerlo para adaptarlo a sus intereses privados. Las medidas o disposiciones intergubernamentales se utilizan con frecuencia para la aplicación de programas públicos nacionales. Así, en los sistemas federales, los organismos federales delegan en los gobiernos regionales y locales la prestación de servicios públicos en sus áreas de jurisdicción. Pero la aplicación final de la intervención pública depende de los burócratas al nivel de calle,²³ las medidas políticas se crean mediante su aplicación, mediante las rutinas que desarrollan y los atajos que encuentran para hacer frente a sus tareas rutinarias, por tanto, su comprensión del programa influye en los resultados del mismo.

f) Respuesta de los administrados.

Siguiendo con el análisis de las tres propiedades de Lundquist (1987), Vedung establece que la comprensión, la capacidad y la predisposición de los destinatarios que participan en la aplicación del programa, influye en el rendimiento y en los resultados. La comprensión de la intervención por parte de los destinatarios puede influir en su comportamiento, por más claro que esté formulado, un programa desconocido por los solicitantes en potencia no contribuirá a los resultados para los que se creó. El conocimiento del programa por parte de los destinatarios se puede completar con comunicados de prensa, anuncios, folletos, programas de radio y televisión; sin embargo, esto puede ser insuficiente para llegar a todos los destinatarios. La adaptación temporal del programa a la situación de decisión del destinatario es un factor de capacidad, susceptible de explicar el resultado. Si el programa llega en un momento formativo puede provocar resultados inmediatos, pero en caso contrario no. En cuanto a la

22 El concepto de captación está vinculado a la teoría de captación regulatoria, la cual sugiere que los organismos o los burócratas se ven arrastrados a actuar de manera favorable para los destinatarios.

23 Término aplicado por Lipsky y del cual hemos hecho referencia en apartados anteriores.

predisposición, el programa puede ser aplicado de manera más eficiente entre algunos interlocutores, debido a la existencia de partidarios fervientes, personas intensamente comprometidas con la causa y dispuestas y capaces de dedicarle todo su tiempo y energía.

La exposición de criterios realizada hasta ahora se ha enfocado hacia el diseño y el sistema de gestión de los programas. Para complementar los criterios de evaluación en un programa hace falta una revisión a la literatura del problema que el programa trata de resolver, esto permitirá obtener algunos criterios útiles al momento de realizar una evaluación de la implementación, además de conocer los supuestos en los cuales se ha basado el programa.

Conclusiones

El libro se centra en presentar herramientas, enfoques y aspectos relevantes que sirvan para valorar las intervenciones y/o acciones públicas que pone en marcha el gobierno, para dar solución a un problema que se considera prioritario. La propuesta ha sido un esquema de evaluación integral con el propósito de mejorar la acción diseñada y ubicar líneas que guíen nuevas investigaciones o cuestiones teóricas, pero, por otra parte, también mejorar la práctica evaluativa. Aunque tradicionalmente el análisis de la política ha estado más enfocado en ofrecer la mejor alternativa para la toma de decisiones, dando a la evaluación un carácter final para establecer si la política ha funcionado o no, si ha cumplido con los objetivos o no, y si es necesario seguir o no con la acción. En la actualidad la evaluación ya no se considera una etapa meramente finalista, y hoy queda más que claro que evaluar involucra a todo el proceso de la política o programa; por lo tanto, se impulsan evaluaciones integrales y no meramente finalistas. En su momento, Bustelo (2001) señaló que la evaluación se ha realizado más con una intención prescriptiva relacionada con la mejora de la política evaluada, más que análisis descriptivos de la fase de evaluación de las políticas, y resaltó la necesidad de más estudios descriptivos sobre los procesos evaluativos (metaevaluaciones).

Esta visión integral se corresponde con la evolución en el concepto de evaluación. Inicialmente enfocada en los objetivos, los resultados y la medición, con carácter a posteriori. Décadas más tarde se resaltó el carácter valorativo de la evaluación, se amplió a todo el proceso del programa (conceptualización y el diseño, implementación y utilidad) y se destacó su fin y funciones. Es su orientación práctica, la cual le otorga a

la evaluación funciones específicas que la distinguen de otras disciplinas: mejora, rendición de cuentas y guiar las decisiones para acciones futuras, aunada la función de ayudar a los consumidores a tomar decisiones sabias (aquí las preguntas evaluativas deben ser de interés práctico para ellos). No obstante, no se descartan funciones más específicas como el control de legalidad, administrativo y financiero, como mecanismo de reforma de la gestión de la administración pública, o como medio que legitima la actuación de los poderes públicos. A pesar de estas funciones, en ocasiones la evaluación se lleva a cabo con propósitos ocultos o menos legítimos, como el encubrir deficiencias o fallos a los superiores, usarla como una actitud de racionalidad, adular un programa débil al justificarlo y elegir deliberadamente solo aspectos que de antemano parecen tener éxito, evitando los aspectos defectuosos, o el ayudar al cliente a crear una imagen positiva del programa con información que asegure el apoyo del público.

Las grandes cantidades de dinero invertidas en los programas y el poco uso de los resultados evaluativos (debido a lo racional de las evaluaciones que no se correspondía con la realidad, a la poca información que existía para realizar buenas evaluaciones, a la entrega tardía de resultados, o resultados que no respondían a las necesidades de los decisores; a los informes extensos y cargados de lenguaje técnico, o simplemente por resistencia al cambio dentro de las organizaciones) centraron la atención en la utilización de los resultados de las evaluaciones. Con base en esto, se ha manifestado como imprescindible que la información responda a las necesidades de las distintas audiencias y poner atención en la difusión de los informes de evaluación, de tal manera que sean amenos y entendibles. Esto implica la participación de los interesados en los resultados de la evaluación durante el proceso de la misma.

A pesar de la resistencia de algunos, la participación de los actores en la evaluación se ve cada vez más como beneficiosa para obtener mejores resultados y hacerla más útil. La comunicación entre el evaluador y las partes interesadas es indispensable para que todos participen en los propósitos y beneficios de la evaluación, así como para negociar los aspectos relevantes de la misma. El trabajo de ambos es complementario: el evaluador es el experto en los métodos y técnicas, pero las partes interesadas tienen la experiencia en el programa que se está evaluando. La tarea del evaluador ya no se limita a la técnica y la metodología, involucra también las relaciones interpersonales que se dan en un contexto político. El conocimiento de la

complejidad política y social de los programas nos acerca más a la realidad y a tratar con mayor eficacia los problemas sociales. Al llevarse a cabo las evaluaciones en organizaciones, la emisión de juicios de valor acerca de lo que se evalúa puede molestar o modificar las relaciones al interior de la organización, por lo tanto la evaluación es altamente política.

La evaluación también debe ser vista dentro del contexto de reformas de la administración pública, aquí sirve como un vínculo para proveer la metodología y/o la información necesitada para implementar las reformas. En Estados Unidos la mayor demanda de evaluaciones ha hecho surgir nuevos enfoques evaluativos y la necesidad de personal calificado, se vivió una etapa de expansión y profesionalización: las universidades comenzaron a ofrecer cursos, aparecieron revistas especializadas y asociaciones profesionales que establecieron normas para su realización. Las evaluaciones dejaron de ser facultad exclusiva de la Federación, los gobiernos locales también las realizan, lo que la ha extendido a todas las áreas y niveles de gobierno. Actualmente sobresale la idea de evaluación como aprendizaje organizacional, es decir, hacer de la evaluación una parte integral de las operaciones diarias de las organizaciones.

Hablar de metodología en evaluación, no se trata solo de la elección entre métodos cuantitativos o cualitativos, sino también de los procesos metódicos que resultan de las posturas tomadas en los niveles ontológico y epistemológico. El entendimiento de los paradigmas (rationales o positivistas, y naturalistas o constructivistas) ayuda a comprender la realidad que los evaluadores y clientes de la misma han percibido en determinado momento y el porqué del uso de ciertos métodos de la evaluación en determinadas épocas, o porqué la importancia de la experimentación o la pluralidad. Hoy se ve cómo la prevalencia del paradigma positivista se está superando y toman fuerza paradigmas alternos que permiten evaluaciones más plurales y que a su vez abarcan más aspectos. La corriente cuantitativa (principalmente) y la cualitativa, han dominado la realización de las evaluaciones. Sin embargo, la elección de uno u otro método depende de las necesidades de la evaluación, donde en ocasiones la combinación de ambos métodos resulta más adecuada. Los efectos no previstos producidos por los programas, amplían el propósito de la evaluación: ya no solo es verificar (cuantitativistas) sino también descubrir cómo el programa podría alcanzar los efectos deseados, aquí el método cualitativo adquirió relevancia.

El decidir qué tipo de evaluación realizar, además de delimitar la evaluación, está condicionado por las preguntas clave a las cuales se quiere dar respuesta, por los objetivos y criterios de valor a utilizar. Esta decisión ayuda a establecer si se realizará, por ejemplo, una evaluación intermedia, externa, de diseño, con el propósito de mejora y/o rendición de cuentas. La teoría de la evaluación resalta los esquemas participativos en los procesos de la evaluación. Estos enfoques implican a las partes interesadas en todo el proceso, desde el establecimiento de las motivaciones detrás de la evaluación hasta la redacción del informe final para hacer factible su utilización.

Cuando se habla de evaluaciones de diseño no solo nos referimos a los objetivos y a la elaboración de los pasos a seguir en el programa, sino también a un estudio del problema a tratar, a sus antecedentes, a un análisis sobre cómo se ha enmarcado el programa respecto de la legislación y políticas nacionales e internacionales. Toda intervención pública debe comenzar con un análisis de la realidad, pues solo mediante la identificación de sus variables (problemas, necesidades, causas y consecuencias) será posible proponer objetivos y medidas que ayuden a superar de manera efectiva la problemática. Además de lo anterior, hay dos enfoques que sobresalen por sus aportes a la evaluación: la Teoría de Cambio del Programa de Carol Weiss (1993) y los estudios de Marcos Interpretativos en políticas públicas de Bustelo y Lombardo (2006) y Viñas (2009). El análisis presentado demostró la gran variedad de criterios que influyen para valorar el diseño de los programas, algunos con énfasis en cuestiones meramente técnicas y otros más en cuestiones de contexto y políticas, pero siempre ambos complementarios. Resulta evidente que en alguno de estos puntos, ciertos aspectos pueden afectar al logro de los objetivos, de ahí la importancia de realizar evaluaciones al diseño de los programas.

Finalmente, la evaluación de la puesta en marcha de un programa involucra tanto la actuación del personal encargado de gestionar el programa, como de las personas a quienes va dirigida la acción. Por evaluación de la gestión del programa se consideran criterios técnicos y administrativos relacionados con la gestión propiamente del programa. Estos criterios condicionarán a su vez la ejecución del programa con la población beneficiaria, por lo que en este tipo de evaluación, además, hay que considerar criterios propios del área que está siendo evaluada.

Referencias

- Acosta, Félix. “Evaluando la política social y el desempeño de gobierno en México: tensiones y dificultades institucionales”, en José María Ramos, José Sosa y Félix Acosta (Coords.). *La evaluación de políticas públicas en México*. México, COLEF e INAP, 2011, pp. 123-146.
- Aguilar Villanueva, Luis. (Ed.). *El estudio de las políticas públicas*. México, Miguel Ángel Porrúa, Grupo Editorial, 2000.
- _____. *La hechura de las políticas públicas*. México, Miguel Ángel Porrúa, Grupo Editorial, 1996.
- _____. *Problemas públicos y agenda de gobierno*. México, Miguel Ángel Porrúa, Grupo Editorial, 2000.
- _____. *La implementación de las políticas públicas*. México, Miguel Ángel Porrúa, Grupo Editorial, 2000.
- Alkin, Marvin. *Debates on evaluation*. Newbury Park, Sage, 1990.
- Alvira, Francisco. *Índices e Indicadores de evaluación de políticas públicas*. Madrid, 1992.
- _____. *Metodología de la evaluación de programas: un enfoque práctico*. Buenos Aires, Lumen/Humanitas, 1997.
- _____. *Metodología de la evaluación de programas*. 3ra ed., Madrid, Cuadernos metodológicos CIS, 2002.
- Amezcuá Viedma, Cesáreo y Antonio Jiménez Lara. *Evaluación de Programas Sociales*. Madrid, Díaz de Santos, S. A., 1996.
- Arellano, David. *et al. Sistemas de Evaluación del Desempeño para organizaciones públicas. ¿Cómo construirlos efectivamente?* México, CIDE, 2012.

- Ballart, Xavier. *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudio de casos*. Madrid, Ministerio para las Administraciones Públicas, 1992.
- _____. “Modelos teóricos para la práctica de la evaluación de programas”, en Joan Subirats y Joaquim Brugué. *Lecturas de Gestión Pública*. Madrid, MAP, 1996, pp. 323-351.
- Bañón, Rafael y Ernesto Carrillo (comps.). *La nueva Administración pública*. Madrid, Alianza Editorial, 1997.
- Bañón, Rafael. *La evaluación de la acción y de las políticas públicas*. Madrid, Díaz de Santos, 2003.
- Bardach, Eugene. *The Implementation Game*. Cambridge, Massachusetts, The MIT Press, 1980.
- Bustelo, María. *La evaluación de las políticas públicas de igualdad de género de los gobiernos central y autonómicos en España: 1995-1999*. Tesis doctoral en Ciencia Política y de la Administración, Madrid, España, Universidad Complutense de Madrid, 2001.
- _____. “¿Qué tiene de específico la evaluación?”, en Rafael Bañón. *La evaluación de la acción y de las políticas públicas*. Madrid, Díaz de Santos, Madrid, 2003, pp. 13-32.
- _____. “La evaluación en el marco del análisis de políticas públicas”, en Roberto Salcedo (Comp.). *Evaluación de políticas públicas*. México, Grupo editorial siglo XXI, 2011, pp. 185-252.
- _____. y Emmanuela Lombardo (eds.). *Políticas de igualdad en España y en Europa*. Madrid, España, Ediciones Cátedra, 2007.
- Campbell, Donald. “Reforms as experiments”, en *American Psychologist*, 24, 1969, pp. 409-429.
- Campbell, Donald. “Grados de libertad y el estudio de casos”, en Thomas Cook y Charles Reichardt (Eds). *Métodos cualitativos y cuantitativos en investigación evaluativa*. 5ta ed., Madrid, Morata, 2005, pp. 80-104.
- Campbell, Donald y Julian Stanley. *Experimental and quasiexperimental designs for research*. Chicago, Rand McNakky, 1966.
- Cardozo, Myriam. *La evaluación de políticas y programas públicos. El caso de los programas de desarrollo social en México*. México, Miguel Ángel Porrúa, 2006.
- _____. “De la evaluación a la reformulación de políticas públicas”, en *Política y Cultura*, 40, México, UAM-Xochimilco, 2013, pp. 123-149.

- _____. (Coord.). *Institucionalización de los procesos de evaluación. Calidad y utilización de sus resultados*. México, UAM-CESOP, 2015.
- Cook, Thomas y Charles Reichardt. *Métodos cualitativos y cuantitativos en investigación evaluativa*. Madrid, Morata, 1986, 2005.
- Cronbach, Lee. “Course improvement through evaluation”, en *Teachers college record*, 64, 1963, pp. 672-683.
- Cronbach, Lee. *Designing evaluations of educational and social programs*. Chicago, Jossey-Bass, 1982.
- Cronbach, Lee. *et al. Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco, Jossey-Bass, 1980.
- Chelmsky, Eleanor. “Comparing and contrasting auditing and evaluation: some notes on their relationships”, en *Evaluation Review*, 9, 1985, pp. 483-508.
- _____. “Evaluating public programs”, en James Perry. *Handbook of public administration*. San Francisco, Jossey-Bass, 1989, pp. 259-273.
- _____ y William Shadish. *Evaluation for the 21st century. A handbook*. SAGE Thousand Oaks, 1997.
- Dye, Thomas. *Understanding public policy*. Englewood Cliffs, Prentice Hall, 1972.
- _____. *What governments do, why they do it, what difference it makes*. Tuscaloosa, Alabama, University of Alabama Press, 1976.
- Elmore, Richard. “Modelos organizativos para el análisis de la implementación de programas sociales”, en Luis Aguilar. *La implementación de las políticas públicas*, México, Porrúa, 2000, pp. 185-249.
- Feinstein, Osvaldo y Gonzalo Hernández. “El papel de la evaluación en México: logros, desafíos y oportunidades”, en Pedro Arizti y Manuel Castro (Eds.). *Mejorando la calidad del gasto público a través del uso de información de desempeño en México*, Washington, D.C., Banco Mundial y SHCP, 2010, pp. 18-22. Consultado en: http://www.coneval.gob.mx/rw/resource/coneval/med_pobreza/Banco_mundial/Libro_Mejorando_la_calidad_del_gasto_publico.pdf.
- Fernández-Ballesteros, Rocío. *Valoración de programas. Introducción a la Evaluación Psicológica II*. Madrid, Pirámide, 1992.
- _____. *Evaluación de Programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid, Síntesis, 2001.

- Filstead, William. “Una experiencia necesaria en evaluación educativa”, en Thomas Cook y Charles Reichardt. *Métodos cualitativos y cuantitativos en investigación educativa*. Madrid, Morata, 2005, pp. 59-78.
- Fink, Arlene. *Evaluation fundamentals. Guiding health programs, research and policy*. SAGE publications, Newbury Park, 1993.
- Fitz-Gibbon, Carol y Lynn Lyons Morris. *How to design a program evaluation*. Beverly Hill, California, SAGE publications, Newbury Park, 1987.
- Fitzpatrick, Jody, James Sanders y Blaine Worthen. *Program evaluation: Alternative approaches and practical guidelines*. 4th ed., Pearson Colleague, 2011.
- Fitzpatrick, Jody. “Origins and current trends in modern program evaluation”, en Jody Fitzpatrick, James Sanders y Blaine Worthen. *Program evaluation: Alternative approaches and practical guidelines*. 4th ed., Pearson Colleague, 2011a.
- _____. “Political, interpersonal, and ethical issues in evaluation”, en Jody Fitzpatrick, James Sanders y Blaine Worthen. *Program evaluation: Alternative approaches and practical guidelines*. 4th ed., Pearson Colleague, 2011b.
- Forns Santacana, María y Juana Gómez Benito. “Evaluación de programas en educación”, en Rocío Fernández-Ballesteros. *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid, Síntesis, 2001, pp. 241-282.
- Freeman, Howard, Peter Rossi y Sonia Wright. *Evaluating social projects in developing countries*. París, OECD, 1979.
- García, Ester. *La evaluación de programas de reforma educativa en España: tres estudios de caso desde un enfoque de metaevaluación*. Tesis doctoral en Ciencia Política y de la Administración, Madrid, España, Universidad Complutense de Madrid, 2003.
- _____. “Metaevaluación de políticas públicas: una visión desde la ciencia política”, en *Revista del CLAD. Reforma y democracia*, 43, 2009, pp. 129-154.
- _____. y Saville Kushner. “Educación, política y políticas de evaluación educativa”, en *Gestión y Análisis de Políticas Públicas*, monográfico sobre evaluación educativa, 3, enero-junio, 2010, pp. 7-15.

- Garde, Juan Antonio (Coord.). *La agencia estatal de evaluación de la calidad de los servicios y de las políticas públicas. Reflexiones y propuestas de creación. Informe Comisión Expertos*. Madrid, INAP, 2005.
- Greene, Jennifer, Holly Kreider y Ellen Mayer. "Combining qualitative and quantitative methods in social inquiry", en Bridget Somekh y Cathy Lewin (Eds.). *Research methods in the social sciences*. London, SAGE publications Ltd., 2005, pp. 274-281.
- Greene, Jennifer. *Mixed methods in social inquiry*. San Francisco, CA, Jossey-Bass, 2007.
- Gomà, Ricard y Joan Subirats (coords.). *Políticas públicas en España. Contenidos, redes de actores y niveles de gobierno*. Barcelona, Editorial Ariel, 1998.
- Guba, Egon. *Toward a methodology of naturalistic inquiry in educational evaluation*. Los Angeles, California, Centre for the study of Evaluation, Universidad de California, L.A, 1978.
- Guba, Egon e Yvonna Lincoln. *Effective Evaluation*. San Francisco, Jossey-Bass, 1981.
- _____. *Naturalistic inquiry*. Beverly Hills, California, SAGE publications, 1985.
- _____. *Fourth generation evaluation*. California, SAGE publications, Newbury Park, 1989.
- Guerrero, Juan Pablo. "La evaluación de políticas públicas", en *Gestión y Política Pública*, IV, 1, 1995, pp. 56-117.
- Ham, Christopher y Michael Hill. *The policy process in the modern capitalist state*. New York St. Martin's Press, 1984.
- Hernández, José Manuel. "Procedimientos de recogida de información en evaluación de programas", en Rocío Fernández-Ballesteros. *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud*. Madrid, Síntesis, 2001, pp. 117-147.
- Hernández Sampieri, Roberto, Carlos Fernández-Collado y Pilar Baptista Lucio. *Metodología de la Investigación*. México, Mc Graw Hill, 2003.
- Hogwood, Brian y Lewis Gunn. *Policy analysis for the real world*. Londres, Oxford University Press, 1984.
- House, Ernest. *Evaluación, ética y poder*. Madrid, Morata, 1994.

- Joint Committee on Standards for Educational Evaluation. *Standards for evaluations of educational programs, projects, and materials*. New York, McGraw-Hill, 1981.
- _____. *Normas de evaluación para programas, proyectos y material educativo*. México, Trillas, 1988.
- _____. *The program evaluation standards. 2nd edition. How to assess evaluations of educational programs*. Thousand Oaks, SAGE, 1994.
- Langbein, Laura. *Discovering whether programs work: A guide to statistical methods for program evaluation*. Santa Monica, California, Goodyear Publisher, 1980.
- Lasswell, Harold. “La orientación hacia las políticas”, en Luis Aguilar. *El Estudio de las Políticas*. México, Porrúa, 2000, pp.79-103.
- _____. “La concepción emergente de las ciencias políticas”, en Luis Aguilar. *El estudio de las Políticas*. México, Porrúa, 2000, pp. 105-118.
- Lindblom, Charles. *El proceso de elaboración de las políticas públicas*. Madrid, MAP, 1991.
- Lipsky, Michael. “Street-Level Bureaucracy and the Analysis of Urban Reform”, en *Urban Affairs Quarterly*, VI, 4, 1971, pp. 391-409.
- _____. y Richard Weatherley. “Street Level Bureaucrats and Institutional Innovation: Implementing Special Education Reforms”, en *Harvard Educational Review*, 47, 2, 1977, pp. 171-197.
- Little, Judith. “Norms of collegiality and experimentation”, en *American Educational Research Journal*, 19, 3, 1982, pp. 325-340.
- Maldonado, Claudia y Gabriela Pérez. *Antología sobre evaluación. La construcción de una disciplina*. México, CIDE, 2015.
- Mejía Lira, José. *La evaluación de la gestión y las políticas públicas*. México, Porrúa, 2003.
- Mény, Yves y Jean-Claude Thoenig. *Las políticas públicas*. Barcelona, Ariel, 1992.
- Monnier, Eric. *Evaluación de la acción de los poderes públicos*. Madrid, Instituto de Estudios Fiscales, 1995.
- Nirenberg, Olga, Josette Brawerman y Violeta Ruiz. *Programación y evaluación de proyectos sociales. Aportes para la racionalidad y la transparencia*. Argentina, Paidós, 2003.
- Norris, Nigel. *Understanding Educational Evaluation*. London, Kogan Page, 1990.

- Olías, Blanca (coord.). *La Nueva Gestión Pública*. Madrid, Prentice Hall, 2001.
- Osuna, José Luis, et al. *Metaevaluación de las políticas evaluadas por EVALÚA DF durante 2009. Informe*, 2011.
- Osuna, José Luis y Carolina Márquez. *Guía para la evaluación de políticas públicas*. Sevilla, Instituto de Desarrollo Regional, Fundación Universitaria, 2000.
- Palumbo, Dennis (ed.). *The politics of program evaluation*. California, SAGE, Newbury Park, CA, 1987.
- Parlett, Malcom y David Hamilton. *Evaluation as illumination: a new approach to the study of innovatory programs*. Occasional Paper. Centre for Research in the Educational Sciences University of Edinburgh, 1972.
- Parsons, Wayne. *Políticas públicas. Una introducción a la teoría y la práctica del análisis de políticas públicas*. México, Flasco, 2007.
- Patton, Michael Quinn. *Qualitative evaluation methods*. Beverly Hills, California, SAGE Publications, 1980.
- _____. *Practical Evaluation*. SAGE Publications, Newbury Park, C. A., 1982.
- _____. *How to use qualitative methods in evaluation*. SAGE Publications, Newbury Park, C. A., 1987.
- _____. *Utilization focused-evaluation. The new century text*. 3er. ed., Thousand, Oaks, SAGE, 1997.
- Posavac, Emil y Raymond Carey. *Program Evaluation, methods and case studies*. Prentice Hall, Englewood Cliffs, 1985.
- Pressman, Jeffrey y Aaron Wildavsky. *Implementation*. Berkley, University of California Press, 1973.
- Ramos, José María, José Sosa y Félix Acosta (Coords.) *La evaluación de políticas públicas en México*. México, COLEF e INAP, 2011.
- Rossi, Peter y Howard Freeman. *Evaluation. A systematic approach*. Londres, SAGE Publications, 1985.
- _____. *Un enfoque sistemático para programas sociales*. México, Trillas, 1989.
- _____. *Evaluation. A systematic approach*. 5th ed., Newbury Park, California, SAGE Publications, 1993.
- Ruano, José Manuel. "El control de gestión en las organizaciones públicas". En Blanca Olías (coord.). *La Nueva Gestión Pública*, Madrid, Prentice Hall, 2001.

- _____. “La doble naturaleza técnica y política de la evaluación del rendimiento”, en Rafael Bañón (coord.). *La evaluación de la acción y de las políticas públicas*. Madrid, Díaz de Santos, 2003.
- Rutman, Leonard. *Planning useful evaluations: evaluability assessment*. Beverly Hills, California, SAGE Publications, 1980.
- Sanders, James. “On mainstreaming evaluation”, en *American Journal of Evaluation*, 23, 3, 2002, pp. 253-259.
- Sanders, William y Sandra Horn. “The Tennessee assessment value-added assessment system (TV AAS): mixed methodology in educational”, en *Journal of personnel evaluation in educational*, 12, 3, 1998, pp. 247-256.
- Scriven, Michael. “The methodology of Evaluation”, en *Perspectives of curriculum Evaluation. Series on Curriculum Evaluation*, 1, 1967, pp. 39-63.
- _____. “An Introduction to Meta-Evaluation”, en *Educational Product Report*, 2, 1969, pp. 36-38.
- _____. “Pros and cons about goal-free evaluations”, en *Evaluation Comments*, 3, 4, 1972, pp. 1-4.
- _____. “Goal-free evaluation”, en Ernest House (ed). *School evaluation: The politics and process*. Berkeley, California, 1973, pp. 319-320.
- _____. “The case study method in social inquiry”, en *Educational Researcher*, 7, 1978.
- _____. *Evaluation thesaurus*. 4th ed., Newbury Park, California, SAGE Publications, 1991.
- _____. *Hard-won lessons in program evaluation*. San Francisco, Jossey-Bass, 1993.
- Shadish, William, Thomas Cook y Laura Levinton (eds.). *Foundation of Program Evaluation. Theories of practice*. Newbury Park, California, SAGE Publications, 1991, 1995.
- Stake, Robert. *Program evaluation: particularly responsive evaluation*. Occasional paper 5, Universidad de Western Michigan, 1975.
- _____. “La evaluación de programas”, en William Dockrell y David Hamilton (Eds). *Nuevas reflexiones sobre investigación educativa*. Madrid, Narcea, 1983, pp. 91-108.
- _____. “Program evaluation, particularly responsive evaluation”, en George Madaus, Michael Scriven, y Daniel Stufflebeam (Eds). *Evaluation models: Viewpoints on educational and human services*

- evaluation*. Massachusetts, USA, Kluwer-Nijhoff Publishing, 1983, pp. 287-310.
- _____. *The art of case study research*. USA, Thousand Oaks, SAGE Publications, 1995.
- _____. *Investigación con estudio de caso*. Madrid, Ediciones Morata, 1998.
- _____. *Evaluación comprensiva y evaluación basada en estándares*. Barcelona, Graó, 2006.
- _____. *Investigación con estudio de caso*. 4ta. ed., Madrid, Morata, 2007.
- Stufflebeam, Daniel. “A depth study of the evaluation requirement”, en *Theory into practice*, 5, 3, 1966, pp. 121-133.
- _____. “The relevance of the CIPP evaluation model for educational accountability”, en *SRIS Quarterly*, 5, 1, 1972.
- _____. “Foundational models for 21st century program evaluation”, en Daniel Stufflebeam, George Madaus y Thomas Kellaghan (eds.). *Evaluation models: Viewpoints on educational and human services evaluation*. 2da. ed., Massachusetts, USA, Kluwer Academic Publishers, 2000, pp. 33-84.
- Stufflebeam, Daniel y Anthony Shinkfield. *Evaluación sistemática: guía teórica y práctica*. Barcelona, Paidós, 1987.
- _____. *Evaluation theory, models, and applications*. San Francisco, Jossey-Bass, 2007.
- Subirats, Joan. *Análisis de políticas públicas y eficacia de la administración*. Madrid, MAP, 1992.
- _____. “Los instrumentos de las políticas, el debate político y el proceso de evaluación”, en *Gestión y Política Pública*, IV, 1, 1995, pp. 5-23.
- _____. “Análisis de políticas públicas y gestión pública: promesas y riesgos de una vida en común”, en Joan Subirats y Quim Brugué. *Lecturas de Gestión Pública*. Madrid, MAP, 1996, pp. 153-162.
- Subirats, Joan, et. al. *Análisis y gestión de políticas públicas*. Barcelona, Ariel, 2008.
- Suchman, Edward. *Evaluative research: Principles and practice in public service and social action programs*. New York, Russell SAGE Publications, 1967.
- Tamayo, Manuel. “El análisis de Políticas Públicas”, en Rafael Bañón y Ernesto Carrillo (eds.). *La nueva administración pública*. Madrid, Alianza, 1997, pp. 281-312.

- _____. *La construcción de un concepto de evaluación de políticas públicas para la democracia*. Tesis doctoral en Ciencias Políticas y de la Administración, Madrid España, Universidad Complutense de Madrid, 2002.
- Tyler, Ralph. “General statement on evaluation”, en *Journal of Education Research*, 35, 1942, pp. 492-501.
- _____. *Basis principles of curriculum and instruction*. Chicago, Chicago Press, 1949.
- _____. *Principios básicos del currículum*. Buenos Aires, Troquel, 1973.
- Valles, Miguel. *Técnicas cualitativas de investigación social: reflexión metodológica y práctica profesional*. Madrid, Síntesis, 2003.
- Vestman, Ove y Ross Conner. “The relationship between evaluation and politics”, en Ian Shaw, Jennifer Green y Melvin Mark (Eds.). *The SAGE handbook of evaluation*, SAGE publications Ltd., 2006. pp. 225-242
- Vedung, Evert. *Evaluación de políticas públicas y programas*. Madrid, Ministerio de Trabajo y Asuntos Sociales, Instituto Nacional de Servicios Sociales, 1997.
- Viñas, Verónica. “Catástrofes y cambio de políticas públicas. Prestige seis años después. Un análisis de marcos interpretativos”, en *Revista Española de Investigaciones Sociológicas*, 127, 2009, pp. 121-157.
- Weiss, Carol. *Evaluation research*. Upper Saddle River, N. J., Prentice Hall, 1972.
- _____. *Using social research in public policymaking*. Lexington, Lexington Books, 1977.
- _____. “The many meanings of research utilization”, en *Public Administration Review*, 39, 5, 1979, pp. 426-431.
- _____. *Investigación evaluativa: métodos para determinar la eficacia de los programas de acción*. México, Trillas, 1987, 1990.
- _____. “Theory-based evaluation: past, present, and future”, en Debra Rog y Deborah Fournier. *Progress and Future Directions in Evaluation. Perspectives on Theory, Practice, and Methods. New Directions for Evaluation*. San Francisco, 1997, 2004, pp. 41-55.
- _____. *Evaluation. Methods for studying programs and policies*. Upper Saddle River, N. J., Prentice Hall, 1998.
- _____. “The interface between evaluation and public policy”, en *Evaluation*, V, 4, 1999, pp. 468-486.

- Wholey, Joseph, John Scanlon, Hugh Duffy, James Fukumoto y Leona Vogt. *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Washington, D.C., USA, The Urban Institute, 1970.
- Wholey, Joseph. *Evaluation: Promise and performance*. Washington, D.C., USA, The Urban Institute, 1979.
- Wholey, Joseph, Harry Hatry y Kathryn Newcomer. *Handbook of program evaluation*. San Francisco, California, Jossey Bass Publishers, 1994.
- Williams, Walter y Richard Elmore (eds.). *Social Program Implementation*. New York, Academic Press, 1976.
- Worthen, Blaine. "Program Evaluation", en Herbert Walberg y Geneva Haertel (Eds.). *The International Encyclopedia of Educational Evaluation*. Oxford, Pergamon Press, 1990, 42-47.
- _____. y James Sanders. *Educational evaluation: Theory and practice*. Wadsworth, New York, 1973.
- _____. *Educational Evaluation: Alternative Approaches and Practical Guidelines*. New York, Longman, 1987.
- Worthen, Blaine, James Sanders y Jody Fitzpatrick. *Program evaluation: alternative approaches and practical guidelines*. New York, Longman, 1997.

Aspectos teóricos de la evaluación de las políticas públicas,
se diseñó en formato digital electrónico en la Dirección de Ediciones
y Publicaciones de la Universidad Autónoma
del Estado de Hidalgo, en el mes de septiembre de 2023.

El gobierno mexicano comienza a aceptar que la evaluación es un proceso que no puede evadir si quiere mejorar las acciones gubernamentales y lograr los resultados planteados. Reconoce también que llevar a cabo evaluaciones le otorga al mismo tiempo legitimidad, pues sirve como mecanismo de rendición de cuentas y apoyo a los tomadores de decisiones para que decidan de manera más informada. En México, la evaluación ha sido una herramienta tradicionalmente no usada ni aprovechada por los gobiernos en la mejora de las políticas y los programas públicos. La evaluación de políticas públicas es, dentro del ciclo de estas, el ámbito menos desarrollado; sin embargo, se hace imprescindible su estudio de una manera clara y sistemática, de modo que pueda contribuir al conocimiento de la misma. Académicamente existe en el país poca referencia al tema; sólo en años recientes, a raíz de la creación de instituciones dedicadas a la evaluación de programas y políticas públicas, comienzan a aparecer publicaciones o seminarios. La práctica evaluativa también ha comenzado a ser más frecuente a partir del año 2007, debido a las reformas hacendarias.

Este libro presenta una revisión de la literatura sobre los temas más relevantes y actuales en evaluación de programas y políticas públicas con el propósito de servir como guía básica a quienes buscan aproximarse al tema. El texto no pretende ser exhaustivo ni profundizar en cada uno de los aspectos presentados, sino acercar al interesado en los temas relevantes y actuales en el ámbito de la evaluación de políticas públicas y programas, de los cuales es preciso conocer para ejercer y mejorar esta práctica. Los temas se abordan a través de una revisión de la literatura, se incluyen las principales aportaciones de los referentes en la evaluación a nivel internacional y en lengua española.



UNIVERSIDAD AUTÓNOMA
DEL ESTADO DE HIDALGO



CONSEJO
EDITORIAL