

# Evaluación y metaevaluación en las políticas y programas públicos

Estado del arte

Myriam Cardozo Brum

Colección Teoría y Análisis



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA  
UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades





EVALUACIÓN Y META-EVALUACIÓN  
EN LAS POLÍTICAS Y PROGRAMAS PÚBLICOS  
ESTADO DEL ARTE

Esta publicación de la División de Ciencias Sociales y Humanidades de la Universidad Autónoma Metropolitana, Unidad Xochimilco, fue dictaminada por pares académicos externos especialistas en el tema.

Primera edición: 7 de marzo de 2012

D.R. © Universidad Autónoma Metropolitana  
UAM-Xochimilco  
Calzada del Hueso 1100  
Col. Villa Quietud, Coyoacán  
C.P. 04960 México, DF.

ISBN:

ISBN de la colección Teoría y análisis: 978-970-31-0929-6

Impreso y hecho en México / Printed and made in Mexico

Evaluación y metaevaluación  
en las políticas y programas públicos  
Estado del arte

Myriam Cardozo Brum



**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades



UNIVERSIDAD AUTÓNOMA METROPOLITANA

Rector general, Enrique Fernández Fassnacht

Secretaria general, Iris Santacruz Fabila

UNIVERSIDAD AUTÓNOMA METROPOLITANA-XOCHIMILCO

Rector, Salvador Vega y León

Secretaria, Beatriz Araceli García Fernández

DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES

Director, Jorge Alsina Valdés y Capote

Secretario académico, Carlos Alfonso Hernández Gómez

Jefe de la sección de publicaciones, Miguel Ángel Hinojosa Carranza

CONSEJO EDITORIAL

José Luis Cepeda Dovala (presidente) / Ramón Alvarado Jiménez

Roberto Constantino Toto / Sofía de la Mora Campos

Arturo Gálvez Medrano / Fernando Sancén Contreras

COMITÉ EDITORIAL

Graciela Lechuga Solís (presidenta)

Francisco Luciano Concheiro Bórquez / Anna María Fernández Poncela

Felipe Gálvez Cancino / Diego Lizarazo Arias

Jaime Sebastián Osorio Urbina / Alberto Isaac Pierdant Rodríguez

José Alberto Sánchez Martínez / Verónica Alvarado Tejeda

Diseño de portada: Irais Hernández Guereca

Asistencia editorial: Varinia Cortés Rodríguez

# Índice

Prólogo .....	9
<i>José L. Osuna</i>	

Introducción .....	15
--------------------	----

## CAPÍTULO I

### Papel de la evaluación en el contexto de la administración pública

La administración pública tradicional .....	21
El estudio de las políticas públicas .....	23
El <i>New Public Management</i> .....	30
Aportes sociológicos .....	37
Enfoques emergentes .....	40
Comentario final .....	42

## CAPÍTULO II

### La evaluación desde el punto de vista teórico

Concepto de evaluación .....	43
Justificación de la evaluación .....	46
Tipos de evaluación .....	48
Actitudes y enfoques .....	54
Contexto político-organizativo de la evaluación .....	61
Comentario final .....	64

CAPÍTULO III  
Métodos y técnicas de evaluación

Debate sobre el método .....	65
Experimentación y métodos cuantitativos .....	67
Métodos cualitativos .....	76
Comentario final .....	82

CAPÍTULO IV  
Experiencias de evaluación  
en organismos internacionales y países desarrollados

Ámbito internacional: ¿control o evaluación participativa? .....	85
Países desarrollados .....	90
<i>Estados Unidos: los dilemas de la experimentación</i> .....	92
<i>Reino Unido: un efectivo trabajo de convencimiento y capacitación</i> .....	95
<i>Canadá: ¿evaluación real o mero ritual?</i> .....	98
<i>Francia: en búsqueda de un modelo adaptado a su idiosincrasia</i> .....	101
<i>España: ¿una formalidad para cumplir con la Unión Europea?</i> .....	114
Evolución global europea .....	115
Comentarios finales .....	117

CAPÍTULO V  
Desarrollos empíricos de la evaluación en América Latina

América Latina: crecimiento lento pero recientemente acelerado .....	119
México: una sistematización largamente postergada .....	132
Profesionalización de la evaluación .....	146
Comentarios finales .....	148

CAPÍTULO VI  
Evaluación de programas sociales en México

Particularidades de la evaluación de políticas sociales .....	152
Evaluaciones de Pronasol .....	154
Evaluaciones de Progresá 1998-2000 .....	159

Auditorías de desempeño de Progresa 1999-2000 .....	163
Evaluación y auditoría de desempeño de Progresa 2001 .....	164
Evaluación de programas federales a partir de 2001 .....	168
<i>La etapa 2001-2005</i> .....	169
<i>El periodo iniciado en 2006</i> .....	172
Evaluación de políticas y programas del Gobierno del Distrito Federal .....	178
Comentarios finales .....	182

## CAPÍTULO VII Metaevaluación

Dificultades para establecer procesos de evaluación .....	183
Conceptos e instrumentos de metaevaluación .....	187
Criterios de metaevaluación .....	196
Tipos de metaevaluación .....	198
Experiencias de metaevaluación fuera de México .....	201
Experiencias mexicanas de metaevaluación .....	205
Comentarios finales .....	211

## CAPÍTULO VIII Ejemplos de metaevaluaciones realizadas en México

Metaevaluación de programas federales de 2002-2003 .....	213
Metaevaluación de programas federales de 2005-2006 .....	219
Metaevaluación de programas del GDF de 2008-2009 .....	222
Comentarios finales .....	229
Consideraciones finales .....	231
Anexo .....	237
Normas para la evaluación de programas del Joint Committee on Standards for Educational Evaluation	
Bibliografía .....	241
Glosario .....	261



## Prólogo

La obra que el lector tiene en sus manos es, desde el propio título y en primer lugar, una consentida licencia editorial que no se corresponde con el riguroso y contrastado contenido de la misma; y sólo cabe interpretarla como un ejercicio de modestia de la autora, coherente con su compromiso con “lo público” desde su intensa dedicación docente e investigadora en la Universidad.

En efecto, siguiendo las acepciones generalmente aceptadas y más conocidas de la expresión, la subtitulación “Estado del arte” implica “ir tras las huellas” del asunto en cuestión –en nuestro caso, la evaluación de políticas públicas–, establecer cómo se trata el tema en la bibliografía, cómo se encuentra en la actualidad y describir las tendencias.

Es cierto que nos encontramos ante un trabajo con una sólida y amplia formulación heurística junto a una rigurosa contrastación hermeneútica que permite a los investigadores interesados, mediante su lectura, disponer de una información solvente y actualizada para colegir de manera sencilla y amena lo que del asunto hay hoy en Europa y América.

Pero, no lo es menos, que –tras las numerosas y aclaratorias citas y las más de 300 referencias bibliográficas y documentales– la obra de Myriam Cardozo, muestra todo su “valor añadido” cuando con un enfoque clásico de la sociología política y de las organizaciones, atiende y entiende el momento actual de la evaluación de políticas públicas no sólo con una visión descriptiva de “lo que hay”; sino también, con un posicionamiento deontológico y metodológico sobre “lo que debe haber”. Y así, avanzando con un característico análisis estructural de lo general a lo particular; de la síntesis y revisión conceptual y metodológica de la/s teoría/as al uso, la obra nos acerca a la aplicación, al estudio de caso.

Aún más de consideración positiva resulta la específica dedicación, un capítulo, a poner de manifiesto la necesidad de la metaevaluación para constatar la calidad de la evaluación; pero, también para cerrar “el bucle” que permite considerarla como herramienta para el aprendizaje, para la formación de los propios gestores de las políticas públicas.

En fin, al respecto de los contenidos y en coherencia con lo dicho de la obra y su autora, es de destacar el carácter propositivo de las consideraciones finales que conjugan con la explicitación de su compromiso social con el que entiende “el más urgente problema nacional” (¿mundial?), confrontación de la realidad económica con el incremento de la pobreza.

Sin embargo, con todo ello, he de confesar que lo dicho no tendría el énfasis aportado si no contase con la que, a mi entender, es la mayor virtud del libro: su oportunidad. Es así, una obra que va a llegar a las librerías y a los fondos bibliotecarios de nuestras universidades no sólo en forma, sino en tiempo y plazo.

En efecto, para quienes pensamos que el momento actual de crisis financiera internacional –“volátil” y hábilmente “voladora” de Oeste a Este del Atlántico-Norte– es tiempo “de política”; y, consecuentemente, entendemos que ésta es una crisis de confianza generalizada que provoca la caída de la demanda como consecuencia de las fuertes e instantáneas limitaciones al crédito que supone la rápida imposición de duras condiciones financieras, el incremento de la incertidumbre en los mercados financieros y la reducción vertiginosa de la riqueza mundial.

Los que entendemos que la lógica de funcionamiento del sistema sigue siendo la consecución de un “monopolio virtual” con base en la mejora de la excelencia en calidad/precio de los bienes y servicios, lo que hace exigible la complicidad público-privada, al ser el conocimiento y la información las nuevas variables de productividad. Y eso es realmente novedoso y revolucionario, porque incide en las relaciones entre el capital y el trabajo provocando el tránsito del llamado “triángulo precario” (definido por *The Economist* en 1964; donde los sindicatos libres, para serlo, aceptaron el compromiso con la estabilidad del sistema, contribuyendo al crecimiento de la producción y la estabilidad de precios; es decir, haciendo aumentar el valor añadido de las empresas con base en la negociación colectiva) al “triángulo del conocimiento” (relatado por Durao Barroso en *El País*, en 2005) que conforman educación, investigación e innovación.

Y aun confesando que no hay recetas mágicas, sí hay caminos lógicos. La academia, con modestia, tiene que opinar lejos de las pantallas mediáticas, sin provocar escándalo innecesario; es decir, hay que buscar las vías de un nuevo contrato social entre el capital y el trabajo en un contexto que se ha manifestado muy obscenamente, en términos de desempleo, de pobreza, de exclusión social y para ello, no puede seguir contribuyendo a justificar el sometimiento del Estado a una dieta de adelgazamiento rápida, instantánea, que provocará más crisis financieras y bursátiles, la persistencia del desempleo, los desequilibrios en la distribución de la renta, la disminución de la cohesión, la continuidad en los éxodos migratorios, los conflictos sociales, las agresiones al medio ambiente... El objetivo es conseguir un Estado que sea cada vez más eficiente y cada vez más transparente; dentro de las ciencias sociales, consolidar la evaluación de políticas públicas como herramienta de la planificación económica.

De ahí la oportunidad del trabajo de Myriam Cardozo. El asunto es más importante de lo que pudiera parecer a primera vista; porque, además de racionalizar la intervención pública, de informar y abrir cauces a la participación ciudadana, de lo que en última instancia se trata es de la ampliación de la democracia.

Ciertamente, los investigadores sociales no hemos encontrado una fórmula mágica para conseguir todo ello; pero sí disponemos de técnicas rigurosas que amplían los resultados de las auditorías “al uso” o de los tradicionales controles que hasta ahora vienen ejerciendo las Cámaras y Tribunales de Cuentas. En primer lugar, porque no se trata tan sólo del necesario control de ejecución y legalidad del gasto público ya realizado. Se trata de intervenir antes, durante y después del proceso, para aumentar su racionalidad y mejorar la eficacia y eficiencia de las políticas, actuando con transparencia, proporcionando información y rindiendo cuentas de lo que se hace con los dineros públicos; y también, porque se requiere la participación ciudadana. Esa eficiencia y esa transparencia son, en este primer cuarto del siglo XXI, las mejores legitimaciones posibles de la intervención del Estado y, combinando adecuadamente el “estado de la cuestión” y el “estado del arte”, no hay posición ideológica sensata que pueda oponerse.

En todo caso, es obligado reconocer que una actividad sistemática como ésta, consistente en valorar la intervención mediante la aplicación rigurosa de procedimientos propios de las ciencias sociales está, obviamente, sujeta al método científico; pero también a los valores predominantes de la sociedad en que tiene lugar el ejercicio de la evaluación.

Así, la obra que está usted leyendo es de gran utilidad en este momento de reconsideraciones, de conocer “por qué” y “para qué” se hacen evaluaciones según las diversas y muy dispares aproximaciones metodológicas, campo en el que la “batalla” va para largo. Los valores que dan sentido profundo a la herramienta como acompañante de la planificación económica no conforman un todo armonioso; es más, con frecuencia entran en conflicto porque la equidad puede ser incompatible con la eficiencia, la libertad con la igualdad o la justicia con la equidad. Es preciso un ejercicio de modestia, porque es aquí donde estriba la enorme dificultad de estimar el valor de las políticas públicas. En efecto, aunque hemos avanzado mucho y tras la propuesta de Christie y Alkin<sup>1</sup> podemos alcanzar un amplio grado de consenso al respecto de que nuestras diferencias están en el mayor o menor énfasis que cada evaluador da a métodos, valores o usos; es justo reconocer que la evaluación no dispone de métodos o herramientas para resolver conflictos de valores que constituyen el núcleo ético-político de la vida social.

Ninguna actividad científica o profesional dispone de esos métodos. Pretender lo contrario equivaldría a tratar de usurpar la vida política a la ciudadanía; “o, aún peor, intentar transformar a los ciudadanos en súbditos, ahora no del rey sino de los nuevos y tecnificados burócratas, científicos, especialistas que no cesan de fomentar la inquietante utopía de la superación de la condición política del ser humano”.<sup>2</sup>

Lo que sí puede hacer la evaluación, de ahí la importancia de su contribución a la sociedad, es poner de manifiesto las implicaciones y consecuencias fundamentales de las políticas públicas en relación con esos valores; mostrando, por ejemplo, en qué medida son más o menos eficientes, suponen una distribución equitativa de riesgos y beneficios, son preferibles a sus alternativas o, hasta qué punto los objetivos previstos pueden, incluso, acentuar las desigualdades existentes. La evaluación no resuelve el conflicto entre objetivos, pero permite conocer sus términos. Y eso ya es empezar a resolverlo.

<sup>1</sup> Christie y Alkin (2004; 2008) a propósito de la metáfora del “árbol de teorías de la evaluación”.

<sup>2</sup> Carlos Román del Río, primer presidente y fundador de la Sociedad Española de Evaluación.

A ello y a conocer las diferentes experiencias de aplicación en los organismos internacionales y en espacios diversos –Canadá, Estados Unidos, Reino Unido, España, Francia y, en especial, América Latina y México– contribuye en gran medida el trabajo ante el que este prefacio ha pretendido situar al lector.

Mi reconocimiento a la autora y a los editores que retoman la idea de hacer una publicación sobre la evaluación de las políticas públicas; convengan conmigo en que resulta difícil, en publicaciones académicas, reconocer tras seis años “un maridaje” tal que les llevó antes, a ser agentes de la palanca para la implementación de la herramienta en México y, ahora, desde el rigor y con un análisis valorativo sobre lo ya hecho, a reflexionar en prospectiva, avizorando su consolidación sobre la base de desarrollos futuros.

Por último, permítanme explicitar el agradecimiento a mi colega, la profesora Myriam Cardozo, por haberme brindado la oportunidad de saludar desde el Prólogo la aparición pública de su trabajo. En estos momentos, cuando muchos de mis compañeros economistas –quizás “indoctrinados por Cambridge”– sólo se sienten pretenciosamente científicos cuando compiten con las ciencias exactas en el uso y el abuso de los modelos econométricos, resulta muy reconfortante ser reconocido como investigador social.

*José L. Osuna*

Catedrático de la Universidad de Sevilla;  
presidente de la Sociedad Española  
de Evaluación de Políticas Públicas



## Introducción

Si bien en algunos países industrializados la evaluación de servicios públicos cuenta con un siglo o más de experiencia (Reino Unido, Estados Unidos), en la mayoría de los países occidentales la evaluación en el campo de las políticas y los programas públicos ha cobrado importancia en las últimas décadas.

América Latina se ha incorporado recientemente a este proceso con algunos pocos casos nacionales que muestran avances importantes, aunque aún insuficientes. En México, las preocupaciones en torno a la evaluación datan de 1917 y se intensifican a finales de la década de 1960; sin embargo, el proceso de evaluación de programas gubernamentales comenzó a instrumentarse de manera sistemática a partir de 1997 con la creación del Programa de Educación, Salud y Alimentación y se generalizó en forma obligatoria para todos los programas gubernamentales sujetos a Reglas de Operación, con base en la decisión incluida en el Presupuesto de Egresos de la Federación, aprobado por la Cámara de Diputados en el año 2000.

Dicho proceso de evaluación consiste en la realización de una investigación aplicada, de tendencia interdisciplinaria, cuyo objetivo es conocer, explicar y valorar el nivel de logros alcanzado (resultados e impactos) por las políticas y programas públicos, así como aportar elementos al proceso de toma de decisiones para mejorar los efectos de la actividad evaluada.

La necesidad de los gobiernos de aumentar su eficacia en la solución de problemas públicos, de hacer un uso más eficiente y transparente de los recursos disponibles y contribuir a la democratización de su gestión, son algunas de las principales razones que explican la relevancia práctica que ha adquirido el tema.

Así, contar con una evaluación integral permite a los gobiernos direccionar y ajustar estratégicamente sus políticas y programas hacia el logro de sus objetivos y metas, a partir de la comprensión de sus resultados o *outputs* (productos o servicios) e impactos o *outcomes* (efectos provocados por los productos o servicios en el problema que se trata de resolver). Para ello resulta necesario conocer las características de los recursos empleados (cantidad, calidad, oportunidad, etcétera) y de los procesos que han permitido su transformación en resultados (productividad, eficiencia, etcétera), así como sus principales condicionantes contextuales (económicas, sociales y políticas).

Múltiples autores han señalado la importancia acrecentada que el proceso de evaluación adquiere en el contexto de los países que, en el marco de su disciplina fiscal, disponen de presupuestos particularmente restringidos de recursos para la atención de los problemas vinculados a la pobreza y el desarrollo social. Esto debido al riesgo de no emplearlos en forma totalmente satisfactoria por falta de información que retroalimente el proceso de gestión.

Otro factor que aumenta su importancia lo constituyen los procesos de reforma del Estado que incluyen la descentralización de funciones para garantizar la eficiencia y efectividad de su gestión y mostrar una equitativa distribución de los esfuerzos.

La evaluación también resulta útil en procesos de democratización caracterizados por el arribo a los cargos de gobierno de políticos provenientes de diversos partidos, con orientaciones y propuestas diferentes, que necesitan asegurar la congruencia de sus acciones. Dicha congruencia puede resultar aún más difícil de lograr cuando se aspira a desarrollar una gestión incluyente de otros actores sociales, (organizaciones de la sociedad civil y ciudadanía en general) como la manifestada en los planes y programas gubernamentales de México.

Finalmente, en países de gran tamaño, la evaluación resulta particularmente relevante para conocer la forma en que se aplican los programas nacionales en los lugares más distantes y de difícil acceso, aun los que cuentan sólo con cierto nivel de desconcentración administrativa, como es el caso de los que se encuentran a cargo de la Secretaría de Desarrollo Social en México.

A pesar de las diversas justificaciones prácticas mencionadas y de la importancia teórica que la evaluación asume en el marco del estudio de las políticas públicas y del *New Public Management*; en la práctica, a menudo se prescinde de realizarla por múltiples razones (falta de recursos, de tiempo, de

consenso sobre sus métodos, su utilidad, etcétera), sus resultados no se difunden o son poco utilizados en la toma de decisiones.

En particular, y a pesar de los esfuerzos realizados en los últimos años para que los informes finales de evaluación de programas resulten accesibles por internet, México cuenta con poca información publicada hasta el presente sobre los procesos emprendidos, especialmente los métodos y técnicas aplicados, la relación costos/calidad de los resultados alcanzados, y el nivel de difusión y utilización de dichos resultados en los procesos decisorios.

Adicionalmente, es importante valorar su calidad mediante un tipo particular de evaluación, denominada metaevaluación. Diversos autores plantean que su objetivo es enjuiciar la calidad y el mérito de lo que hacen los evaluadores profesionales, que persigue el perfeccionamiento de sus trabajos y promueve una mejor comprensión de los mismos. Esto implica que la metaevaluación toma como objeto de análisis al proceso de evaluación aplicado a los programas sociales: sus insumos, procedimientos, métodos, resultados e impactos.

Consideramos que el cumplimiento del objetivo anterior exige también incluir y valorar el sistema en que se organizan y desarrollan las actividades realizadas por los evaluadores, incluyendo el papel que desempeñan otros actores o instancias que también inciden en la calidad del proceso de evaluación desarrollado y del informe final logrado. Se trata de factores de su contexto como: condiciones políticas, legales, culturales, y sobre todo, la organización y operación de las instancias que participan en el proceso.

Con esta adición, la metaevaluación analiza y valora la corrección de los métodos de evaluación, sus técnicas operativas y la calidad de los resultados alcanzados, pero también un conjunto de factores contextuales que afectan a los anteriores, con énfasis en la organización y funcionamiento práctico del sistema que conforma el contexto inmediato de la evaluación.

Los informes revisados se analizan con base en la elaboración de listas de chequeo, cuyo estudio se complementa y amplía mediante la aplicación de entrevistas, fundamentalmente en relación con los procedimientos administrativos relacionados con el proceso de evaluación; y con el objetivo de conocer las opiniones de los actores entrevistados sobre los principales problemas enfrentados, los factores que los explican y sus posibles soluciones.

Todos estos indicadores fueron valorados con base en criterios generales de investigación (validez, confiabilidad, representatividad, consistencia, objetividad,

etcétera) y específicos de la metaevaluación (pertinencia de la evaluación, autonomía del evaluador, legitimidad y credibilidad del proceso, calidad y utilidad de los resultados).

La calidad de las evaluaciones ha sido medida a partir de la elaboración de un indicador compuesto de los tres elementos, que recupera las características más importantes del caso federal (rigurosidad en el diseño, recolección y tratamiento de datos; cumplimiento de los requerimientos incluidos en los términos de referencia; y explicitación de un juicio valorativo bien fundamentado a la gestión del programa) y del desarrollado en el gobierno del Distrito Federal (también rigurosidad en el diseño, recolección y tratamiento de datos; manejo solvente del marco teórico-conceptual, jurídico y de planeación global; y nuevamente, explicitación de un juicio valorativo bien fundamentado a la gestión del programa).

Con base en lo anterior, el libro que se presenta<sup>1</sup> ha pretendido poner a disposición de los lectores una síntesis de los conocimientos teóricos y empíricos disponibles en la materia (conceptos, enfoques, métodos y experiencias de evaluación y metaevaluación), así como un breve recorrido por la evaluación en el contexto mexicano y una valoración de su calidad. Su contenido se desarrolla en los siguientes capítulos:

El primero ubica el papel desempeñado por la evaluación en el devenir de la administración pública y los enfoques que la han permeado: burocracia tradicional, estudio de las políticas públicas, *New Public Management*, aportes sociológicos y debates recientes.

<sup>1</sup> Un antecedente importante del presente libro lo constituye el publicado por la autora en 2006 en la Colección “Conocer para decidir” de la H. Cámara de Diputados y Editorial Miguel Ángel Porrúa, con el título *La evaluación de políticas y programas públicos. El caso de los programas de desarrollo social en México*. El mismo se organizaba en dos partes: una relativa a los asuntos conceptuales, teóricos, metodológicos y empíricos de la evaluación y metaevaluación; y otra que informaba sobre las políticas y programas sociales mexicanos en curso, los primeros esfuerzos de evaluación en el país, el marco jurídico vigente, y llevaba a cabo la metaevaluación de las evaluaciones realizadas hasta 2003. El libro que ahora se presenta recupera y actualiza la primera parte del anterior y sólo retoma de la segunda aquello que continúa vigente, agregando un capítulo en el que se ejemplifican las metaevaluaciones realizadas a partir de la investigación presentada en el anterior.

El capítulo siguiente sintetiza los principales desarrollos teóricos en materia de evaluación: conceptos, clasificaciones, enfoques e importancia del contexto político-organizativo.

El tercero es un capítulo de carácter metodológico; discute la pertinencia que para la evaluación tienen los diferentes métodos desarrollados en ciencias sociales, sus técnicas y criterios específicos. Analiza la utilidad de los métodos experimentales, cuasiexperimentales y no experimentales, así como los de carácter cuantitativo y cualitativo.

A continuación se dedican tres capítulos a relevar algunas experiencias nacionales e internacionales en la materia: el primero recoge la situación en organismos multilaterales y países desarrollados; el segundo se refiere a América Latina y ofrece un esquema general de la evolución del caso mexicano; mientras el tercero profundiza en las evaluaciones realizadas en nuestro país, a partir de las experiencias de los programas Pronasol y Progresá en los años 1993-2001, tanto a nivel federal como del gobierno del Distrito Federal.

Los desarrollos teóricos y empíricos disponibles sobre metaevaluación de la calidad de las evaluaciones realizadas constituyen el objeto del capítulo VII; mientras el último presenta algunos ejemplos de metaevaluaciones realizadas.

Finalmente, se sintetizan los principales hallazgos realizados y se enfatizan algunas propuestas para mejorar la situación de la evaluación y metaevaluación, y coadyuvar a la resolución de los problemas planteados con la intención de que México aproveche la enorme oportunidad que tiene actualmente de mejorar la calidad de los trabajos de evaluación e integrar sus resultados a la gestión pública, como forma de incidir más efectivamente en el alivio del más urgente problema nacional: el incremento de la pobreza registrado en los últimos años.



# CAPÍTULO I

## Papel de la evaluación en el contexto de la administración pública

Durante el siglo XX la evaluación fue teóricamente reconocida como una de las funciones fundamentales que integran los procesos de la administración pública; sin embargo, fue recién en la década de 1980 cuando su necesidad práctica parece haberse impuesto en la mayoría de los países.

Para comprender su evolución será necesario revisar los distintos enfoques que han caracterizado a la administración pública, partiendo de la burocracia tradicional, resaltando la importancia que adquirió en el contexto de las políticas públicas y con el advenimiento del New Public Management, así como las aportaciones realizadas por la escuela sociológica y los debates actuales.

Este capítulo pondrá especial atención a la evaluación de resultados de la aplicación de políticas y programas públicos.

### **La administración pública tradicional**

La administración pública, como práctica, puede ser identificada desde el desarrollo de las primeras civilizaciones y especialmente con los grandes imperios. Ya en la época moderna, Woodrow Wilson (1887) la definió como el gobierno en acción, o sea, en términos de R. Uvalle,<sup>1</sup> el conjunto de decisiones y acciones

<sup>1</sup> Notas de la conferencia “Institucionalidad y administración pública”, presentada por el doctor Ricardo Uvalle Berrones, en el Seminario obligatorio de doctorantes de la Facultad de Ciencias Políticas y Sociales, Universidad Nacional Autónoma de México, el 27 de mayo de 2003.

complejas, orientadas a la obtención de resultados, que concretan la capacidad de gobernar para la persecución manifiesta del bienestar social y la implícita de protección al gobierno mismo.

Esta cara visible del gobierno, asumió rasgos de tipo burocrático durante la mayor parte del siglo XX, a menudo sensiblemente diferentes del tipo ideal weberiano (especialmente en América Latina). La administración pública burocrática se caracterizó por una clara distinción entre medios de administración públicos y privados, los primeros se dirigían a la satisfacción de las necesidades colectivas, a partir de servicios públicos. Se la describe como basada en: formación profesional, criterios objetivos, división del trabajo, estructuras jerárquicas, apego a leyes y reglamentos detallados, especificación minuciosa de funciones y actividades, autoridad y lealtad, carrera administrativa, desempeño neutral, impersonal, autorización detallada de gastos, orientado al servicio público, con control central y directo y promoción como único incentivo.

En la segunda mitad del siglo pasado, la burocracia fue generalmente identificada con sus rasgos más negativos. De esta forma, la administración pública tradicional fue acusada de: excesivo crecimiento, centralismo, subutilización de su estructura formal, poca coordinación, incapacidad para atender las demandas, ingobernabilidad, lentitud, ritualismo y rigidez en los procedimientos, apego irrestricto a la norma, la rutina y el formalismo, exceso de control, insuficiente motivación de servicio público, inutilidad y duplicación de trámites, desinformación y pérdida de expedientes, poca racionalidad, autoritarismo, falta de sensibilidad ante los temas empresariales, corrupción, etcétera.

A principios de la década de 1970 dicha percepción negativa se incrementó, influenciada por las ideas de diversos economistas y revistas del campo gerencial, que responsabilizaron a la burocracia de una presupuestación tradicional y una administración financieras ineficientes para garantizar el logro de las metas macroeconómicas y detener el crecimiento del sector público (Barzelay, 2001; Crozier, 1974).

En la década de 1980, caracterizada por profundas crisis (fiscal, de endeudamiento, de bienestar, etcétera) y cambios acelerados (de tecnología, condiciones de trabajo, etcétera), la administración pública tuvo que comenzar a cambiar su imagen y hasta su nombre. Así, el término de gestión pública (retomado

de la tradición franco-canadiense: *gestion publique*)<sup>2</sup> ha sustituido al desgastado vocablo de “administración pública”, que carga con contenidos poco deseables para el discurso de la Modernización o Reforma del Estado. A menudo este concepto aparece confundido con el de gerencia pública (traducción habitual de *public management*), que se refiere sólo a la función directiva de este proceso,<sup>3</sup> con énfasis en la persecución de la eficiencia y preocupada por asuntos ligados a liderazgo y capacidad de mando.

La administración tradicional descrita no dio lugar a desarrollos importantes en materia de evaluación, puesto que su interés estuvo centrado en el control concomitante, directo y detallado de las actividades, en el marco del derecho administrativo y los reglamentos específicos. Sin embargo, en paralelo al proceso reseñado, se había ido gestando un enfoque innovador, que revisaremos a continuación, en el que la evaluación ocupa un papel importante.

### **El estudio de las políticas públicas**

En la segunda mitad del siglo XX comenzó a destacar el estudio de las políticas (*policies*) públicas. Éstas persiguen la resolución de problemas públicos, tomando en cuenta aspectos políticos y también de orden técnico. Se trata de un proceso en el que se entremezclan “las iniciativas de las organizaciones sociales y las de las instancias gubernamentales” (Aguilar, 2000a:8) y que intenta responder a una exigencia de racionalidad, legalidad y eficiencia de los gobiernos, sin necesidad de “comprometerse unilateralmente con alguna específica preferencia conceptual y valorativa sobre el Estado y la acción gubernamental” (Aguilar, 2000a:9). Para Quade (1975) “no es sino un complemento, muy poderoso, del buen juicio, la intuición y la experiencia de los ‘decision makers’” (Quade, 1975:71).

<sup>2</sup> Véanse al respecto las publicaciones periódicas de la OCDE como *Optique. Le bulletin de la gestion publique* [[www.oecd.org/puma/index-fr.htm](http://www.oecd.org/puma/index-fr.htm)]. También documentos del gobierno de Québec, Canadá, publicados en *Modernisation de la gestion publique* [[www.tresor.gouv.qc.ca](http://www.tresor.gouv.qc.ca)].

<sup>3</sup> Paramés (1978:79) la define como una función de alto nivel de todas las organizaciones, un proceso de utilización sabia de personas y cosas, y un grupo de personas responsables del gobierno de grupos sociales de muy diversa condición. Guerrero (1998:20) define a la gerencia pública como el nexo entre las *policies* y la administración.

Una extensa descripción de lo que el concepto puede implicar se encuentra en Subirats (1989, citado en Aguilar, 2000c:22-23): un campo de actividad gubernamental, un propósito general a realizar, una situación deseada, una propuesta de acción específica, las normas relativas a cada problemática, el conjunto de objetivos y programas de acción del gobierno en un campo de cuestiones; o bien, el producto y resultado de una específica actividad gubernamental, el comportamiento gubernamental de hecho, el impacto real de la actividad gubernamental, el modelo teórico o las técnicas aplicables que sustentan una iniciativa gubernamental, etcétera. La lista demuestra la enorme ambigüedad del concepto y la ausencia de un consenso entre la comunidad de estudiosos que garantice su rigor científico.

En el presente trabajo se define a las políticas públicas como un fenómeno social, administrativo y político específico, resultado de un proceso de sucesivas tomas de posición, que se concretan en un conjunto de decisiones, acciones u omisiones, asumidas fundamentalmente por los gobiernos, mismas que traducen, en un lugar y periodo determinado, la respuesta preponderante del mismo frente a los problemas públicos vividos por la sociedad civil.<sup>4</sup>

Es así que los gobiernos continúan siendo considerados como los principales actores, únicos que disponen de recursos suficientes para ofrecer bienes y servicios que satisfagan las amplias necesidades sociales. Esto no significa que se niegue importancia a la participación de las organizaciones civiles y sociales o los ciudadanos en general, sino que ésta, si bien se estima deseable, no se considera un elemento imprescindible para definir una política como pública.

En relación con la definición propuesta es necesario aclarar que se habla de un proceso de sucesivas tomas de posición, donde se concatenan planes, programas, proyectos, acciones u omisiones, por lo que se descarta que una decisión aislada, específica, constituya una política; que se hace referencia a problemas públicos como aquellos que, por su relevancia para importantes sectores de la sociedad, no son factibles de resolverse en el ámbito privado; y se considera que los gobiernos no actúan de manera absolutamente coordinada, racional y consistente, por lo que es posible encontrar posiciones diferentes entre las organizaciones que lo componen, y en consecuencia, sólo se pueden identificar tendencias predominantes.

<sup>4</sup> La síntesis presentada, incluida en Cardozo (2001) retoma ampliamente elementos desarrollados en los trabajos de O'Donnell y Oszlak (1976) y Oszlak (1980).

Corresponde señalar que este enfoque de análisis también se ha aplicado a un nivel más desagregado de estudios, el que se refiere a los programas públicos, que permiten una mayor concreción de las políticas en relación a sectores y/o temas transversales específicos. Si bien en el pensamiento anglosajón los términos de políticas y programas pueden aparecer casi como sinónimos (Muller, 1994), en el presente trabajo, se entiende que un programa define con precisión sus objetivos operativos, los medios para alcanzarlos, así como las acciones a realizar en un cierto tiempo; mientras que la política constituye “un conjunto complejo de programas, procedimientos y regulaciones que concurren a un mismo objetivo general” (Kessler *et al.*, 1998:1; traducción libre), un sistema concatenado y sinérgico de los mismos, que requiere del desarrollo de importantes procesos de coordinación y articulación para lograr el cumplimiento de su misión.

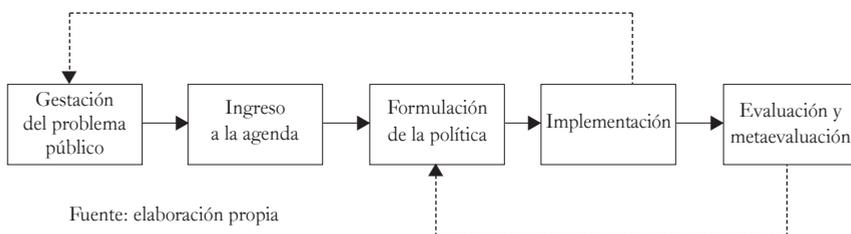
El proceso de políticas puede desagregarse en fases o etapas, enfoque que, aunque no goza de pleno consenso (Dunn, 1981 o Lindblom, 1980; citados en Aguilar, 2000c), resulta dominante entre los estudios de políticas públicas. Esta desagregación ha sido útil para el análisis del proceso, siempre que ha quedado claro que no debe entenderse como una secuencia temporal estricta, sino como una aproximación meramente analítica.

El número de fases del denominado ciclo de las políticas varía con los autores (Laswell, 1971; May y Wildawsky, 1977, citado en Aguilar, 2000b; Brewer y DeLeon, 1983; Anderson, 2000, entre otros), pero todos coinciden en incluir las etapas de aplicación o implementación, evaluación y terminación. El esquema más comúnmente empleado en nuestro medio se conforma de:

- La *formulación*, que incluye un diagnóstico y definición del problema por parte del organismo estatal, mismo que condicionará el conjunto de alternativas estratégicas de solución que comenzará a generar y evaluar *a priori* para seleccionar la que considere más apropiada. Cabe aclarar que el diseño de políticas, como ya se mencionó, suele estructurarse bajo la forma de programas más concretos, y aún de proyectos específicos, para operacionalizar su ejecución en la siguiente etapa.
- La *implementación*, que traduce en decisiones operativas y acciones concretas la política antes decidida, de forma de impactar el problema social que le dio origen, procediendo al control o seguimiento de lo formulado.

- Transcurrido el tiempo necesario para poder medir los resultados e impactos alcanzados, podrá procederse a realizar una *evaluación a posteriori* de la política para describir el proceso que se llevó a cabo, explicar sus determinantes y condicionantes y calificar su grado de efectividad. La evaluación permitirá decidir si la política continúa aplicándose y si lo hará con o sin modificaciones, pudiendo dar lugar a la repetición de todas o algunas de las fases señaladas (Diagrama 1).

DIAGRAMA 1  
*Proceso de políticas y programas públicos*



En diferentes contextos y periodos históricos, el peso relativo del estudio de políticas se inclinó hacia alguna de las etapas señaladas. En el caso de los Estados Unidos, en la década de 1960, el énfasis estuvo en su iniciación; posteriormente, la evaluación resaltó las ineficiencias de las políticas en curso y se volvió predominante. Su análisis identificó serios problemas de implementación, con lo que esta etapa cobró importancia en la segunda mitad de la década de 1970, acompañada en los casos de mayor fracaso, por la consideración de la etapa de terminación. En la década de 1980, se produjo el redimensionamiento del Estado y se regresó al interés por el diseño de políticas, capaces de aumentar la eficiencia de la gestión pública (Aguilar, 2000b). Hoy asistimos a una renovación del interés por la etapa de evaluación en el contexto de las reformas que revisaremos en el punto siguiente.

Por otra parte, las diferentes ciencias que participan del enfoque inter y multidisciplinario de las políticas públicas, también han concentrado su interés en el análisis de algunas fases en detrimento de las restantes:

Es así que los politólogos han preferido estudiar la formación de la agenda de gobierno, los administradores públicos la implementación y los economistas se han dedicado enteramente a la construcción y selección de las opciones de política (Aguilar, 2000c:19).

Podríamos agregar también que los sociólogos se interesan por la gestación de los problemas en la sociedad, aunque en los últimos años, en Francia, han sustituido a los economistas en la realización de evaluaciones. Estas especializaciones han logrado importantes aportaciones al conocimiento, pero también han tenido resultados no deseados como un economicismo que poco recupera los análisis de impacto social y político en los procesos de formulación de políticas o un fuerte abandono de la evaluación, objeto principal de preocupación en el presente trabajo.

Laswell (1951), iniciador de este campo de estudios, definió a las ciencias de las políticas como las “disciplinas que se ocupan de explicar los procesos de elaboración y ejecución de las políticas” (citado en Aguilar, 2000a:47), con base científica, interdisciplinaria y al servicio de gobiernos democráticos. En un trabajo posterior (1971) agregaría que era posible diferenciar “el conocimiento ‘del’ proceso de decisión y ‘en’ el proceso de decisión” de políticas, refiriéndose a las habilidades profesionales necesarias para participar en la toma de decisiones públicas y las habilidades científicas requeridas para contribuir a la invención de una teoría y práctica confiables (Laswell, 1971:117).

La diferenciación anterior daría lugar a dos corrientes: por un lado, quienes se autoidentificaron fundamentalmente bajo el rótulo de *Policy Analysis*, interesados exclusivamente en el momento normativo-técnico de análisis y diseño de la política, desde un enfoque unidisciplinario; y, por otro, los partidarios de las *Policy Sciences*, preocupados sólo por entender los procesos decisivos, como suceden en los hechos, y las acciones puestas en marcha, desde una perspectiva multi o interdisciplinaria.

En este trabajo se ubica a la evaluación de políticas y programas públicos como parte del conocimiento “del” proceso, en la medida en que busca analizar científicamente los resultados de los procesos de decisión e implementación. La misma puede ser desarrollada por el propio gobierno o el Congreso, pero también por organizaciones civiles y universidades que representen los intereses de la sociedad.

El tratamiento de las políticas públicas aparece también surcado por diferentes enfoques que podríamos esquematizar de la siguiente manera: un enfoque de racionalidad total, que supone un decisor individual o diversos decisores organizacionales que actúan como bloque unificado, disponen de información completa y una capacidad ilimitada de cálculo que les permite revisar exhaustivamente todas las alternativas disponibles y encontrar siempre la mejor solución a los problemas públicos (optimización). Este enfoque se caracteriza por ser normativo y fuertemente cuantitativo.<sup>5</sup>

Un segundo acercamiento desde el análisis del proceso organizacional que pone el énfasis en la explicación de la decisión tomada con base en las rutinas y procedimientos estándares vigentes en las organizaciones. Parte de la idea de una racionalidad limitada que lleva a alcanzar decisiones satisfactorias, en lugar de óptimas.<sup>6</sup>

Existe otro enfoque fundamentado en el proceso político incrementalista en que cada actor se mueve por sus propios intereses y negocia con los demás sin compartir sus fines, por lo que se considera que la optimización no es posible (no se cuenta con información completa ni existe capacidad ilimitada de cálculo) ni deseable (no existen objetivos compartidos).<sup>7</sup>

Un cuarto enfoque aporta el conjunto de elementos no racionales del proceso decisorio, en que se recuperan elementos como intuición y experiencia desde perspectivas psicosociológicas, antropológicas y filosóficas.<sup>8</sup>

Finalmente podemos encontrar mezclas de los distintos enfoques, como el planteado por A. Etzioni, quien critica al modelo racional por las razones ya planteadas, pero también al incremental por su tendencia conservadora, que descarta los cambios radicales. Propone una recuperación selectiva de elementos de los distintos enfoques anteriores.

<sup>5</sup> Sus orígenes los podemos encontrar desde la propuesta cartesiana, desarrollada en Teoría de Juegos por Von Neumann y Morgenstein y en Teoría de la Estrategia por Von Clausewitz.

<sup>6</sup> El enfoque es iniciado con los trabajos conductistas de Herbert Simon, *El comportamiento administrativo* (original de 1947, nueva edición de 1972) y *Models of Man, social and rational* (1957), y continuado por Cyert, March, Olsen, Cohen, entre otros.

<sup>7</sup> Charles Lindblom retoma la idea de incrementalismo planteada por el enfoque anterior, atribuyéndole una valoración diferente en su famoso artículo "The science of muddling trough", aparecido en 1959.

<sup>8</sup> Citado a pie de página por Allison (1971) y desarrollado en Steinbruner (1974).

En el presente trabajo de evaluación *a posteriori* optamos por esta última posición al considerar que las políticas y programas públicos atraviesan y provocan diversos procesos, con distintos tipos de racionalidad no exhaustiva: 1) políticos, en la definición de los valores y objetivos a perseguir para alcanzar algún nivel de satisfacción de la ciudadanía que permita el mantenimiento o incremento de la legitimidad de los actores sociales (incluido el gobierno); 2) técnicos, de ajuste de medios a fines en la identificación y selección de políticas y programas a aplicar para el logro de los anteriores (racionalidad administrativa instrumental, de tipo limitada); y 3) sociales, vinculados a la posible participación de la ciudadanía y las organizaciones de la sociedad civil en los procesos previos y al imprescindible análisis de los impactos provocados en la población por las estrategias aplicadas.

También se ha cuestionado si el estudio de las políticas públicas constituye una disciplina o un campo de acción multidisciplinario (muy vinculada con la misma discusión que enfrenta la administración pública como disciplina), científico o no científico. Dada la trascendencia de esta discusión, en la que no es posible profundizar en este trabajo, se le considerará un campo profesional en el que concurren múltiples disciplinas, como sucede con la medicina. Se pueden hacer evaluaciones que apliquen el método científico para describir y explicar las políticas implementadas y sus efectos. También es posible que la formulación se base en diagnósticos científicos de la situación, pero el análisis de toma de decisiones está fundado en la previsión de posibles consecuencias futuras, marcadas por la incertidumbre, donde las categorías de “verdadero” y “falso” pierden sentido y se sustituyen por otras como “probable” o “factible”, en cuyo trabajo de prospectiva influyen también aspectos como experiencia, sensibilidad, intuición, etcétera, además del conocimiento científico, lo que dificulta la definición de su estatus.

Actualmente, Majone (1996) identifica tres rasgos relativamente nuevos que caracterizan las preocupaciones de las políticas públicas: el redescubrimiento de la necesidad de eficiencia, el reconocimiento de la importancia estratégica de la credibilidad política, y el aumento en la delegación de poderes a unidades independientes para la hechura de políticas; todos ellos estrechamente vinculados con el tema de la evaluación.

### El *New Public Management*

En las dos últimas décadas, y con la intención de resolver múltiples problemas burocráticos, dio inicio un esfuerzo de reforma del Estado y modernización de la administración pública, impulsado por el enfoque denominado *New Public Management* (NPM), que ha dado un nuevo y fuerte impulso a los procesos de evaluación.

Este enfoque, traducido con dificultades en nuestro medio como Nueva gestión o Nueva gerencia pública, se apoya en el neoinstitucionalismo económico y el *Public Choice*, así como en experiencias prácticas calificadas como exitosas en varios países sajones, particularmente en Nueva Zelanda, Reino Unido y Australia, que introdujeron reformas presupuestales dirigidas al control del gasto y el logro de resultados, acompañadas de importantes modificaciones legales y cambios institucionales, así como de un fuerte proceso de privatización y de rendición de cuentas.<sup>9</sup> Se trata de un movimiento que ha recibido un decidido apoyo de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), el Fondo Monetario Internacional (FMI) y el Banco Mundial (BM).

Resulta importante subrayar la idea de que gestión y gerencia pública son campos preexistentes al nacimiento del NPM y que, a su interior, existen enfoques distintos y hasta opuestos a éste en algunos aspectos, como el defendido por Bernardo Kliksberg (1989 y 1994) que no admite la crítica generalizada a la ineficiencia del sector público y lejos de tender a equiparar la administración pública con la privada, enfatiza sus diferencias, incluso las de la gerencia social frente a la gerencia económica, con su particular papel de lo político, la preocupación por la articulación (no subordinación) de las políticas económicas

<sup>9</sup> Guerrero (1999:129) plantea que Michel Messenet, director general de la Función Pública en Francia, empleaba ya, en 1975, terminología e ideas del NPM como orientación al mercado, privatización, descentralización, preocupación por los costos y la eficiencia administrativa, consideración del ciudadano como cliente, etcétera, en su libro *La Nouvelle Gestion Publique: pour un État sans bureaucratie*, Editions Hommes et Techniques, París. A su vez, Messenet reconoce haberse inspirado en los trabajos de un administrativista privado, Octave Gelinier, quien hizo el prólogo de su libro y sería, a su vez, autor de *Le secret des structures compétitives*, publicado en 1967. En el contexto del Reino Unido, Guerrero señala que fue recién en 1993 que apareció una obra similar, *Public sector management*, de N. Flynn.

y sociales, los procesos de negociación, la priorización de los intereses de las grandes mayorías empobrecidas, la aplicación del enfoque de políticas públicas, el énfasis en la participación como mecanismos de control social y el desarrollo de programas sociales masivos.

El NPM no constituye un nuevo paradigma, ya que al respecto no existe consenso entre los científicos del campo (Kuhn, 1971), ni un programa sistemático de investigación sobre el mismo (Lakatos, 1989), ya que no contiene un conocimiento ordenado y sistematizado en la materia, sino una combinación de estudios relacionados, con una orientación común: el incremento de la eficacia, economía, eficiencia (lema de la triple E utilizado por Margaret Thatcher) y calidad del servicio. El mismo ha sido definido como un campo de discusión sobre las intervenciones del gobierno en materia de políticas, tanto en términos de análisis sistemático como de gerencia práctica, enfocada a los procesos políticos y organizacionales así como también a sus contenidos sustantivos (Barzelay, 2001); como un movimiento genérico que cruza diferentes sectores de las políticas y fronteras nacionales (Hood y Jakson, 1997); o se ha reconocido a su interior un conjunto de modelos diferenciados como el manejo de la eficiencia, la reducción y descentralización del Estado, la búsqueda de la excelencia, o la orientación al servicio público (Ferlie *et al.*, 1996).

Entre las diversas estrategias que el NPM propone para lograr los objetivos de su agenda internacional encontramos: reducción del tamaño de la burocracia y selección de lo que el gobierno debe hacer, descentralización, separación entre la formulación y la implementación de la política, jerarquías aplanadas, facultación y responsabilización, coordinación de las interdependencias, competencia entre organizaciones públicas, generación de valor para el cliente, cooperación público-privado, enfoque de gobernanza participativa (de consumidores, empresarios y empleados públicos), distinción entre el financiamiento y la prestación de servicios públicos, apertura a la competencia, privatización, utilización de técnicas de gerencia privada, flexibilización, simplificación de trámites, desregulación, democratización, enfoque proactivo de problemas, énfasis en costos, calidad, libertad de elección,<sup>10</sup> satisfacción y *empoderamiento* del cliente, cultura orientada a resultados, focalización de programas, evaluación

<sup>10</sup> Basada en la noción de “soberanía del consumidor” de Ludwig Von Mises.

del desempeño, utilización de incentivos de mercado, transparencia contable, rendición de cuentas, etcétera.

Para Abate (2000) su aplicación requiere de una evolución general de la sociedad que garantice: 1) una amplia base profesional suficiente para gerenciar en forma autónoma la gestión pública (competencias y capacidades de iniciativa de los agentes), 2) disposición de redes informáticas y de tratamiento de datos, y 3) capacidad de los sistemas de información de medir los costos y los productos de las acciones de la gestión para evaluar su calidad y su impacto en la sociedad.

En el contexto latinoamericano, el Consejo Científico del CLAD, aprobó en octubre de 1998 el documento “Una nueva gestión pública para América Latina”, que explicitó su posición oficial sobre la reforma gerencial. Asumió que este tipo de reformas ofrecía las mejores posibilidades en términos de democracia y progreso para la región, pero también planteaba el desafío de compatibilizar las tendencias mundiales con sus especificidades (necesidad de consolidar la democracia, de recuperar el crecimiento económico y de reducir la desigualdad social). Dicho documento propuso: profesionalización de la alta burocracia, transparencia de la administración pública, descentralización de la ejecución de servicios, desconcentración de las actividades del gobierno central, orientación a resultados, mayor autonomía compensada con nuevas formas de control, distinción entre actividades exclusivas del Estado y otras del espacio público no estatal, orientación de los servicios hacia el ciudadano-usuario, democratización del poder público y mayor responsabilización de políticas y funcionarios. Para su puesta en práctica el CLAD señalaba como prerequisite completar la reforma burocrática weberiana y redefinir las relaciones entre el Estado y la sociedad, construyendo una esfera pública que involucrara a esta última en la gestión de las políticas públicas.

Uvalle (1998, 2001, 2002 y 2003) coincide con el CLAD en que las reformas deben tener dos objetivos: gestión democrática y gobierno eficiente. El último, que caracteriza como un “gobierno facilitador”, se define por una gestión conjunta con los ciudadanos para la atención de los problemas comunes y por el control democrático del rendimiento institucional. Considera que la recuperación del espacio público por la sociedad reduce las posibilidades de volver a contar con “Estados magnificados”.

Más recientemente, Serra y otros (2007; citado en Cardozo, 2009) retoman la propuesta del NPM para impulsar la denominada Gestión para resultados, que promueve mayor racionalidad en la toma de decisiones con objeto de alcanzar

eficacia, eficiencia y economía (las tres E) y trata de coadyuvar a la comprensión de los factores y procesos que afectan el éxito de un sistema de gestión. De esta manera busca contribuir al logro de los tres objetivos principales de la gestión pública: *a)* asegurar la constante optimización del uso de los recursos públicos en la producción y distribución, *b)* lograr un proceso de producción de bienes y servicios públicos que mejore su productividad y sea transparente, equitativo y controlable, y *c)* promover la efectividad de los organismos públicos, mediante un mejor desempeño de sus dirigentes y trabajadores.

En relación con los resultados alcanzados por las reformas, un informe producido por el Comité de Gerencia Pública de la OCDE (1995) señalaba que era muy pronto para juzgar el impacto de los cambios realizados, pero también se reconocía una lamentable escasez de evaluaciones rigurosas de los mismos que permitiría construir mayor confianza y promover una cultura basada en el enfoque gerencial. Se aseguraba que las experiencias “sugerían” éxito en el control del gasto público y que se disponía de “evidencia considerable” en el caso de agencias individuales para “hacer más con menos”, incrementar su eficiencia y la calidad de sus servicios. En cambio, se reconocía la dificultad y complejidad de medir la efectividad de las reformas introducidas, por lo que muchas conclusiones se calificaban de “tentativas” y exigían continuar manteniendo como supuestos del enfoque los referidos a que la utilización de incentivos permitirá incrementar los resultados o que la descentralización mejorará el nivel de logros al reducir los costos asociados con los controles centrales.

Nueva Zelanda, el caso emblemático del NPM, tampoco cuenta con evaluaciones consistentes sobre el éxito de sus reformas: si bien logró éxitos financieros y macroeconómicos de gran magnitud, éstos aparecen acompañados de algunos resultados no deseados como un fuerte incremento en el desempleo.<sup>11</sup>

<sup>11</sup> Aumento del crecimiento del PIB del 77%, del PIB per cápita de 160%, de la productividad del trabajo del 400%, disminución de la inflación del 57%, del déficit en cuenta corriente del 24%, etcétera, en el periodo de 1983-1997, comparado con el de 1973-1983, aunque éstos no han permitido recuperar los niveles de los mismos indicadores que el país gozó en el periodo 1963-1973. Los anteriores se acompañaron de un aumento de la tasa de desempleo del 267% (pasó de 1.8 a 6.6%), en la comparación de los mismos periodos y una disminución del 17% del total de empleos, en el sector público, en 1989-1997.

En el caso de América Latina (AL), Cohen (2001) plantea que, a pesar de la importancia de los mecanismos de descentralización, externalización y focalización implementados en los programas sociales como parte de las reformas en curso, no existe claridad sobre los beneficios de su aplicación. Trabajos como el de Spink (2001) o el de Longo y Echebarría (2001) argumentan a favor, en términos de posibilidades, pero también en contra, recogiendo las limitaciones del enfoque.

Esta falta de pruebas contundentes ha llevado a que los estudiosos del NPM se dividan en un grupo que defiende teóricamente las bondades de la propuesta (Barzelay, 1992 y 2001; Osborne y Gabler, 1994; Uvalle, varios años; Osborne y Plastrik, 1998; Hood, 1994; OCDE, 1995) y otro que la cuestiona por su ideología, inconsistencias o pretensión de aplicación universal (Pollit, 1993; Christensen y Laegreid, 2001; Guerrero, 2004; Brugué, 2004). Thoenig (2002) señala que la marcha triunfal del NPM ha comenzado a frenarse en el contexto del BM, el FMI, la OCDE, Reino Unido, Nueva Zelanda, etcétera.

En el presente trabajo, más que involucrarnos en la polémica anterior, nos interesa resaltar el nuevo impulso que el NPM ha dado a la evaluación como herramienta de buena gestión. La OCDE (1995) señalaba cinco características fundamentales a desarrollar por los gobiernos en la siguiente década; dos se referían a la evaluación continua de la efectividad de las políticas y reformas y al impulso de un enfoque de gobernanza, más participativo.

Es que más allá de la necesidad que históricamente ha tenido la administración pública centralizada de contar con retroalimentación sobre sus logros para corregir y rediseñar sus acciones, los cambios introducidos en materia de descentralización, la delegación de mayores facultades, flexibilidad y libertad (disminución del control jerárquico) a los gerentes para el logro de los objetivos comprometidos, la participación de agencias no gubernamentales, etcétera, incrementan el requerimiento de una cultura orientada a resultados<sup>12</sup> que necesita ser evaluada, tanto para la aplicación de incentivos positivos y negativos, como para la transparencia contable y la rendición de cuentas, insumos básicos de la gobernanza participativa. Así, la traducción de los objetivos estratégicos del

<sup>12</sup> Esta orientación hacia el logro de resultados fue retomada de antecedentes como el Planning-Programming-Budgeting System (PPBS) de Estados Unidos, la Rationalisation des Choix Budgetaires (RCB) francés o el enfoque de administración por objetivos.

gobierno en metas operativas y la verificación de su cumplimiento constituyen la verdadera esencia de la gerencia pública (OCDE, 1995). También la Unión Europea (UE) y la Organización de las Naciones Unidas (ONU, 1994), menos permeables al enfoque del NPM, han insistido en la necesidad de la evaluación de los programas y el desempeño de los funcionarios.

Por las razones anteriores, el NPM refuerza los sistemas de información y evaluación de los resultados de la gestión pública (políticas, programas, organizaciones, funcionarios), focalizándolos hacia la medición de la eficiencia y economía de operación, la calidad de los servicios y las relaciones costo/efectividad. En algunos países como Australia los resultados se miden en conjunto, pero en otros como Nueva Zelanda o el Reino Unido se hace una clara diferenciación entre resultados en tanto productos o servicios brindados (*outputs*) y sus impactos (*outcomes*), siendo los primeros, responsabilidad de los gerentes, y los segundos, de los ministros. Bélgica, Francia o España han enfatizado el análisis de la calidad de los servicios.

Los sistemas de evaluación requieren de dos actividades complementarias: 1) el diseño de indicadores apropiados de seguimiento continuo del desempeño y sus correspondientes estándares (con base en logros previos, logros de otras organizaciones similares, etcétera), y 2) la realización de evaluaciones periódicas en profundidad. La primera, también denominada monitoreo, ofrece señales sobre la posible existencia de problemas y la segunda se aplica para su verificación, explicación y valoración, o sea, intenta establecer relaciones causa-efecto entre los factores explicativos y el nivel de logro alcanzado por el programa, mismos que permitan tomar decisiones sobre su continuación, conclusión o modificación en aspectos de financiamiento, delegación de facultades y/o rediseño de sus contenidos. De esta forma las actividades de evaluación se integran con las de planeación-presupuestación, dando lugar a una gerencia por resultados.

[Dicha evaluación] no es una ciencia exacta [...] este punto no siempre ha sido entendido por los legisladores e incluso por los ministros [...] quienes pueden hacer un uso acríptico de la información sobre resultados en la creencia de que todo logro puede ser plena y precisamente cuantificado (OCDE, 1995:104, traducción libre).

Dicha información debe ser considerada en el marco de su definición conceptual, metodológica y empírica y puede haber enfrentado dificultades de distintos tipos: la definición poco operativa de los objetivos planeados, la multiplicidad y conflicto entre objetivos, la complejidad de la información analizada, la difícil verificación de relaciones de causalidad, la incompatibilidad de tiempos para su realización y su utilización en el ciclo presupuestal, la insuficiencia de recursos debida a las reducciones del gasto público, la falta de independencia de los evaluadores, etcétera.

En el mismo sentido, Abate (2000) señala la dificultad de gerenciar los servicios no comerciales del Estado, en pos de la eficacia en la producción de servicios colectivos para beneficio de la sociedad, en comparación con la simple búsqueda de rentabilidad en una empresa privada. En cuanto a la evaluación de la gestión por resultados la considera una disciplina a crear, pues se requiere de nuevas formas de pilotaje capaces de manejar la multiplicidad de factores que determinan un cierto nivel de logro, muchos de ellos relativos a la sociedad y no sólo provocados por las acciones de la gestión pública. Duran y Thoenig (1996) agregan otro nuevo factor a esta complejidad: las características multiorganizacionales de los programas públicos, que obligan a la evaluación a superar los límites de las organizaciones individuales como unidad de análisis.

A pesar de las dificultades anteriores, la información sobre logros debe ser usada en los procesos decisorios; para ello debe ser reportada a las personas que la requieren y en forma diferente según sus responsabilidades (ministros, jefes ejecutivos, legisladores, etcétera). “[L]a obtención de ganancias de logro reposa sobre la comprensión por los actores de los procesos en curso y sobre las laxitudes que les son dadas para actuar” (Abate, 2000:53, traducción libre). Además, el autor señala que la orientación a resultados no se logrará plenamente si la evaluación se limita a alimentar el interior de la gestión pública; considera fundamental el diálogo con los ciudadanos y el parlamento como un medio poderoso para estimular su logro.

Por otra parte, la calidad de la evaluación de resultados puede ser mejorada mediante la realización de metaevaluaciones, o sea, “evaluaciones de las evaluaciones”.

## Aportes sociológicos

En los estudios clásicos de ciencias políticas, el Estado y el gobierno se definen por la dominación y el poder que ejercen sobre la sociedad y la economía en un cierto territorio; en cambio, en la sociología de la acción pública se definen como un proceso de coordinación de actores, grupos sociales e instituciones para alcanzar los fines propios, discutidos y definidos colectivamente en contextos fragmentados e inciertos (Mayntz, citada en Commailles y Jobert, 1998; Norman, 2007). Este proceso resulta más cercano al concepto de *governance* (traducido en español como gobernanza o gobernancia), o capacidad para gobernar, vista a partir de la producción de redes de actores gubernamentales y cívicos en el desarrollo de políticas públicas. En su estudio se retoman las ideas planteadas por el neoinstitucionalismo sociológico y la teoría de la contingencia, en torno a la importancia de la cultura y el marco institucional de las organizaciones para explicar las dificultades que enfrentan los procesos de cambio, especialmente los inducidos.

La sociología política maneja el concepto de regulación social para referirse al modo de coordinación de las diferentes actividades o relaciones entre actores, la asignación de recursos relacionada con estas actividades y actores, la estructuración de los conflictos, su prevención y resolución; así como los procesos de creación y aplicación de reglas para asegurar la coexistencia, en un mismo territorio, de individuos con intereses divergentes, incluso contradictorios y potencialmente conflictivos. Se vincula al tema de la integración social y trata de la acción colectiva organizada, bajo formas multipolares, poco o no formalizadas, y no necesariamente coercitivas. La diversidad de la última provoca que su coherencia sea problemática, sus formas variadas y cambiantes y su efectividad incierta; lo que obliga a pensar en formas renovadas de intervención social.

Para esta disciplina, la regulación social se efectúa mediante tres mecanismos complementarios:

- Regulación por el Estado, en que éste estructura los conflictos, distribuye recursos y coordina actividades y grupos, para lo que requiere capacidad de control y sanción.
- Regulación por el mercado, que permite la organización de intercambios en el capitalismo.

- Regulación por la cooperación o reciprocidad, basado en valores, normas, identidades, intercambios movidos por la confianza, solidaridad entre miembros de la familia, comunidad, etcétera (Commailles y Jobert, 1998).

La sociología de las organizaciones, por su parte, ha superado el énfasis casi exclusivo en la organización formal, para recuperar el análisis de la informalidad, ya planteado por Simon (1972). La organización se ha vuelto un sistema social en funcionamiento concreto, producto de un trabajo de construcción de todos sus miembros, en que lo formal y lo informal se han podido conciliar mediante el estudio del concepto de poder. Se trata de analizar a la organización para identificar “sus actores, intereses, recursos, restricciones, comportamientos, en el contexto específico de los juegos relacionales que estructuran la organización” (Commailles y Jobert, 1998:39-40, traducción libre), relativizando sus posiciones y poderes específicos, al mismo tiempo que descubriendo las reglas implícitas (de segundo orden o no escritas) que compatibilizan intereses heterogéneos y son respetadas por los actores. La racionalidad de las decisiones individuales se descubre después de describir y explicar el juego colectivo.

Este enfoque de las organizaciones se ha extendido a la comprensión de los sistemas de acción, o sea, las relaciones de interdependencia e intercambio entre individuos y grupos que intervienen alrededor de un problema, lo que nos permite aplicarlo al análisis de las políticas y programas públicos, sustituyendo el enfoque *top down* de la sociología del Estado todopoderoso por un acercamiento *bottom up*.

También aporta el concepto y funcionamiento de las estructuras cognitivas de Bourdieu (1993), que permiten describir representaciones, percepciones y visiones del mundo de los individuos, así como las estructuras que determinan su formación y transformación (Chauviré y Fontaine, 2003). Las estructuras cognitivas, a veces denominadas estructuras mentales, hacen que los individuos acepten el mundo como algo natural, provocando una especie de “sumisión” cognitiva en lugar de una verdadera actividad cognitiva. Constituyen la prolongación, a nivel individual, de las estructuras objetivas de los diferentes “campos” sociales.

La aplicación de las aportaciones de la sociología de las organizaciones ha sido muy productiva en el contexto francés, en que el Estado ha dejado de disponer de importantes recursos financieros y jurídicos, ya no controla los programas que, si bien son decididos por el poder central, son aplicados por las administraciones locales. Ahora ya no se trata de producir infraestructura o de brindar servicios

públicos, sino de resolver problemas difusos, poco estructurados, a los que no se adaptan los mecanismos burocráticos. En este contexto, la coordinación y la negociación se consideran fundamentales para la articulación de fines colectivos a partir de intereses de múltiples actores diferenciados y se impone la cooperación y comunicación en el marco de redes de solución de problemas.

Un ejemplo importante de esta aportación sociológica lo constituye el “modelo de las regulaciones cruzadas”, desarrollado para comprender el cambio de los asuntos públicos franceses en las décadas de 1960-1970, centrándose en el análisis empírico de las estrategias de los actores, interpretando los juegos colectivos a los que responden las anteriores y los modos de regulación que gobiernan estos juegos y constituyen las características más pertinentes de los sistemas, que conforman redes más o menos organizadas de relaciones humanas (Crozier y Thoenig, 1975).

Más recientemente, la sociología de las organizaciones ha orientado el estudio de la gestión pública territorial a partir del modelo de institucionalización de la acción colectiva (Duran y Thoenig, 1996; Duran, 1999; y Roggero, 2005 y 2006), en el marco de las transformaciones sufridas por el Estado francés (debilitamiento de su poder, descentralización, aparición de actores locales y grupos de interés, multiplicación de las externalidades, etcétera). Su aplicación a nivel territorial permite descubrir esquemas de cooperación informal que explican la solidaridad entre actores que ocupan puestos, niveles jerárquicos e inserciones institucionales diferentes (Thoenig, 1994).

En materia de evaluación este enfoque sugiere analizar: 1) la forma en que diversos actores con intereses divergentes logran volverlos compatibles en la organización, a pesar de todas sus contradicciones y conflictos, 2) las normas implícitas que regulan sus relaciones recíprocas, 3) las consecuencias que estos arreglos inducen, y 4) el grado en que dichas consecuencias son “pagadas” por grupos internos a la organización social o externalizadas sobre terceros. Esta última manifestación permite superar el corte de los estudios exclusivamente volcados hacia el interior de las organizaciones, al analizar la forma en que los actores externos afectan o son afectados por éstas. También recomienda tomar en cuenta las diferencias culturales<sup>13</sup> en los procesos de adaptación de tecnología

<sup>13</sup> Una caracterización de la diversidad de las culturas nacionales ha sido realizada por Geert Hofstède, en *Cultures and organizations. Software on the mind*, Mc Graw Hill Internacional,

administrativa desarrollada fuera del país (en materia de evaluación, por ejemplo, los estudios experimentales producidos en Estados Unidos).

### **Enfoques emergentes**

Además de las críticas ya mencionadas que se han venido formulando al NPM durante las dos décadas de su aplicación, en la actualidad parece percibirse una mayor tendencia a proponer modelos alternativos o, al menos, ajustes importantes en su aplicación.

Brugué (junio de 2004), después de criticar como debilidad principal del NPM su sesgo hacia la competencia y el mercado, mismo que considera que lejos de transformar a la administración pública tradicional, ha reproducido sus rasgos eficientistas, propone una alternativa que denomina “administración pública deliberativa”, que se percibe muy cercana al gobierno por discusión sugerido por Guerrero (1999). Dicho modelo, sin abandonar la eficiencia como objetivo instrumental, se enfocaría al objetivo sustantivo de la modernización administrativa en una forma cercana a la propuesta sociológica: resolver los conflictos y responder a las demandas del entorno social. Esto exige desarrollar una administración basada en la confianza y el diálogo, para lo cual se requieren nuevas formas y estilos de liderazgo.

En el caso de América Latina y en los últimos años, se percibe un viraje importante en la posición del CLAD respecto de las reformas inspiradas en el NPM. Su Consejo Directivo convocó, en 2003, a la Quinta Conferencia Iberoamericana de ministros de administración pública y reforma del Estado para debatir el tema de los consensos políticos y sociales necesarios, por considerar que muchas decisiones de reforma “han sido impuestas por urgencias” que han impedido su construcción, lo que dio lugar al consenso de Santa Cruz.

En 2004, el documento “Democracia en América Latina: hacia una democracia de ciudadanos y ciudadanas” del Proyecto de Naciones Unidas para el Desarrollo

---

1991, con base en cuatro dimensiones: desigualdad social o índice de distancia al poder, grado de individualismo-colectivismo, rechazo-aceptación de la incertidumbre, diferenciación y valores asociados al género.

(PNUD) confirmó algunos datos anticipados en 2003 por Latinbarómetro en el sentido del limitado avance de la democracia en la región y el debilitamiento de la gobernabilidad. En el mismo tenor, el documento introductorio a la Sexta Conferencia Iberoamericana de Ministros de Administración Pública y Reforma del Estado convocada por el CLAD, señala que:

[...] gran parte de las condiciones negativas para la gobernabilidad democrática provienen de que las reformas estructurales que hasta ahora se diseñaron y ejecutaron [...] han sido orientadas por las necesidades de un determinado enfoque económico, cuyo resultado ha sido la disminución del apoyo a la democracia (Fernández Toro, junio de 2004:1).

Frente a esta situación, el CLAD propuso redefinir las jerarquías entre economía y política para dar prioridad a la democracia frente al mercado: vigencia del estado de derecho, efectividad de los derechos políticos, civiles y sociales y participación efectiva de la ciudadanía en las decisiones sobre las cuestiones sustantivas de la sociedad. Esto se refuerza en Bresser Pereira y otros (2004), quienes enfatizan la necesidad de que la ciudadanía participe como un actor político clave y se produzca un mayor control social sobre los procesos de reforma.

Más recientemente, este organismo insistió en que:

[...] el objetivo más general de la administración pública es el desarrollo sustentable, considerado en términos económicos, políticos, sociales y ambientales. Se debe poner especial énfasis en la búsqueda de la ciudadanía social, de modo que la gestión pública de mejor calidad debe estar presente en las políticas de educación, salud, seguridad pública, combate a la pobreza y la desigualdad social [...] es en dichas áreas sociales en donde el Estado debe invertir sus recursos financieros y sus competencias administrativas más valiosas (CLAD, 2010:20).

Para ello propuso que la gestión pública continúe valorándose por indicadores de economía de recursos, eficacia y eficiencia; pero, también por la ética, la capacidad emprendedora, el equilibrio entre las acciones gubernamentales, y el empoderamiento ciudadano.

En lo que respecta a la evaluación, Zaltsman (junio de 2004) concluye, con base en el análisis del trabajo desarrollado por Cunill y Ospina (2003), que

en AL hay casos nacionales en que la evaluación de resultados de la gestión resultó muy afectada por los valores y principios del NPM, pero también otros en que la participación de la ciudadanía y la profundización de la democracia representativa desempeñó un rol igualmente importante. Además, no identifica indicios de incompatibilidad o conflicto que impidan a la evaluación diseñada la consideración de valores diferentes como la democracia y la equidad social.

### **Comentario final**

Después de 50 años de relativo desinterés de la burocracia tradicional por la evaluación, ésta fue ganando terreno con el desarrollo de los estudios de políticas públicas e incrementando su importancia en el marco de las estrategias promovidas por el NPM. Los enfoques sociológicos y emergentes tampoco han cuestionado a la evaluación misma, sólo han tratado de infundirle valores y objetivos diferentes a los impulsados por el *New Public Management*.

El siguiente capítulo profundizará en el abanico de aproximaciones teóricas y conceptuales presentes en el desarrollo actual de la evaluación de políticas y programas públicos.

## CAPÍTULO II

### La evaluación desde el punto de vista teórico

Después de esbozar la evolución del papel desempeñado por la evaluación en la administración pública, el presente capítulo tiene por objetivo sintetizar las diversas contribuciones teóricas que a la fecha se han realizado en materia de evaluación de políticas y programas públicos, especialmente en su enfoque *a posteriori*, ofreciendo un panorama general de los conceptos, objetivos, enfoques, actitudes y destinatarios de la misma, y resaltando la importancia del contexto político-organizativo en que se desarrolla.

#### **Concepto de evaluación**

Como tantos otros conceptos de las ciencias sociales, podemos encontrar casi tantas definiciones de evaluación como autores han escrito sobre ella. Sus diferencias suelen girar alrededor de énfasis distintos atribuidos a los elementos que componen el concepto en estudio, por lo que podemos hablar fundamentalmente de un conjunto de definiciones que se complementan, más que de concepciones opuestas. M. Weber habría sido uno de los primeros en definir a la evaluación como “la apreciación (*bewertung*) práctica de un fenómeno sobre el cual nuestra actividad puede ejercer una influencia al adoptar respecto del mismo una actitud de aprobación o desaprobación (citado en Duran y Monnier, 1992, traducción libre).

Sintetizando las aportaciones de múltiples autores (Guba y Lincoln, 1981; Nioche, 1982; Stufflebeam y Shinkfield, 1987; Guba, 1987; Rossi y Freeman, 1993; Mény y Thoenig, 1995; Dye, 1995, citado en IDR, 2000; Weiss, 1998; Duran, 1999; BM, 2004a), es posible considerar a la evaluación como una investigación aplicada, de tendencia interdisciplinaria, realizada mediante la aplicación de un método sistemático, cuyo objetivo es conocer, explicar y valorar una realidad, así como aportar elementos al proceso de toma de decisiones, que permitan mejorar los efectos de la actividad evaluada. Se resalta la importancia atribuida a la emisión de un juicio valorativo sobre dicha actividad, lo que también es compartido por Scriven (1967) y Aguilar y Ander-Egg (1992).

Algunos autores consideran (Monnier, 1995; Solarte, 2002; Díaz y Rosales, 2003) que definiciones como la presentada enfatizan su carácter instrumental, por lo que ensayan definiciones de corte más sociológico y político, referidas a la evaluación como un proceso social e histórico de construcción de consensos y acuerdos, transformación de las ideas y los conocimientos, basado en el diálogo, con enfoques metodológicos múltiples y distintos del discurso racional de la teoría de las decisiones, que contribuye a modificar los programas públicos en forma lenta y marginal. Resaltan la importancia de que la evaluación coadyuve al logro de una democracia efectiva, a partir de la rendición de cuentas del gobierno a los ciudadanos y de su posible participación en el proceso.

En el mismo sentido, desde la antropología social, Parlett y Hamilton (1985) han venido desarrollando una concepción denominada “iluminativa”, que no pretende medir efectos, ni emitir juicios de valor, sino limitarse a describirlos e interpretarlos con metodología cualitativa y enfoque holístico, poniendo énfasis en los procesos organizacionales y de implementación, así como en los mecanismos de participación social. Ha sido especialmente utilizada en la evaluación de programas educativos.

Por otra parte, resulta importante diferenciar a la evaluación del mero control o fiscalización de la gestión, que se enfoca a la utilización de los recursos presupuestales, y de la auditoría tradicional, centrada en verificar la adecuación de los procedimientos legales y administrativos. Zapico (2005, citado en Cardozo, 2009) propone que no sólo se persiga la legalidad del gasto sino que el énfasis se ponga en el aumento de la productividad, e incluso que la atención se traslade del ámbito de los recursos al de los resultados e impactos.

Por otro lado, Cunill y Ospina aportan una nueva diferenciación entre los conceptos de monitoreo<sup>1</sup> y de evaluación.

[El primero es considerado como] la valoración recurrente del avance de una intervención, usando indicadores de desempeño seleccionados *a priori*. Los indicadores pueden ser cualitativos o cuantitativos, y pueden enfocar el proceso y/o resultados. Sin embargo, el seguimiento no explica las causas de lo que se documenta, ni ofrece la oportunidad de aprehender la complejidad de los procesos detrás de las mediciones realizadas [Mientras que la segunda constituye] una valoración puntual de una intervención, mediante el uso de procedimientos y herramientas analíticas para obtener información pertinente, comparable y confiable del proceso, los resultados e impactos, o de los tres, a fin no sólo de documentar los patrones, sino también de explicarlos y/o comprenderlos, capturando su complejidad (Cunill y Ospina, 2008, citadas en Cardozo 2009: 26-27).

Finalmente, la realización de una evaluación implica un conjunto de actividades, cuya secuencia varía según el tipo de evaluación a realizar:

- a) Identificar los efectos provocados por la acción, programa, política o evaluación a evaluar y los costos incurridos.
- b) Aplicar una escala de medición a los efectos identificados (nominal, ordinal, de intervalo o razón).
- c) Comparar la medida lograda con otra que sirva de parámetro para la evaluación (ej.: la medida de efectos logrados en periodos previos por la misma organización, política o programa; en el periodo actual por organizaciones, políticas o programas similares; la determinada en los planes organizacionales como meta esperada, etcétera), fundamentalmente, en el caso de los programas sociales, para constatar si hubo mejoramiento de las condiciones de bienestar de la población.
- d) Explicar los resultados comparativos encontrados, en función del diseño, la implementación, y las condiciones de aplicación del programa.
- e) Emitir un juicio de valor que califique, en conjunto, las actividades realizadas, los servicios brindados, sus efectos y su impacto global.

<sup>1</sup> Si bien monitoreo es el término más utilizado por organismos como el Banco Mundial o el Banco Interamericano de Desarrollo, en el contexto latinoamericano se le conoce más como control o seguimiento.

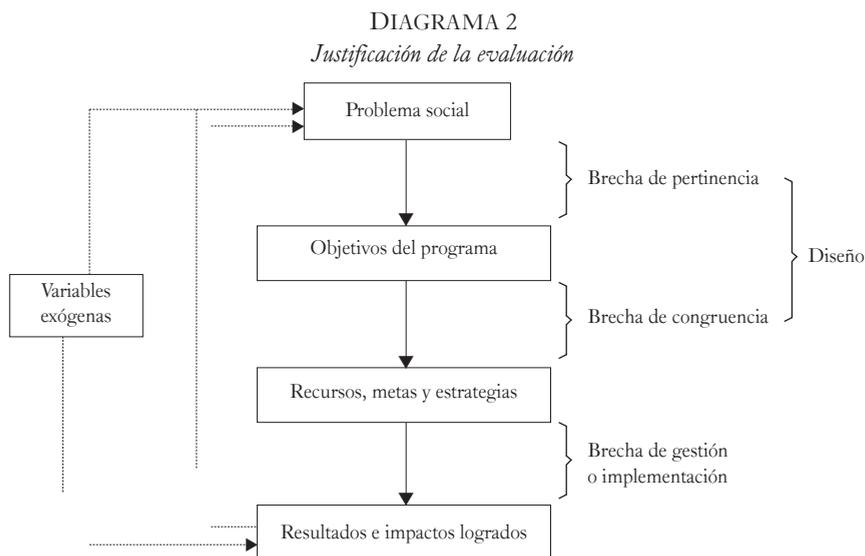
f) Efectuar las recomendaciones necesarias para enfrentar los problemas detectados y aprovechar las fortalezas del programa, de manera de contribuir a un mayor logro de sus objetivos.

Las acciones evaluadas pueden vincularse a diversos objetos de evaluación: personas, organizaciones, políticas, planes, programas, proyectos o evaluaciones. En el presente trabajo el objeto de la evaluación lo constituirán las evaluaciones aplicadas a las políticas y programas públicos, especialmente los que persiguen objetivos sociales.

Los anteriores pueden ser muy diversos (combate a la pobreza, educación, vivienda, salud, medioambiente, empleo, servicios básicos, etcétera), e implicar conocimientos específicos de campos variados.

### Justificación de la evaluación

Tal cual puede verse en el Diagrama 2, cualquier problema social es afectado por múltiples causas contextuales y por el o los programas que se hayan diseñado para tratar de solucionarlo.



Fuente: elaboración propia.

El diseño anterior define fundamentalmente objetivos, metas respecto de resultados y de impactos esperados, estrategias e instancias responsables. Pero sucede que en este proceso pueden generarse al menos tres brechas o diferencias importantes: 1) la brecha de pertinencia del programa y sus objetivos a las necesidades insatisfechas identificadas como problema, 2) la brecha de congruencia entre dichos objetivos, los recursos, las metas correspondientes y las políticas, programas y estrategias definidos para alcanzarlos; y 3) la brecha de gestión o implementación del programa que introduce aspectos como la estructura, los procesos y la cultura organizacional de las instancias participantes, así como su sistema de interrelaciones.

Como consecuencia del proceso descrito se obtiene un cierto nivel de resultados e impactos sobre el problema original, que normalmente difiere del planeado debido a la presencia de las tres brechas señaladas, a las que se suma la posible incidencia de variables exógenas tanto en el problema inicial como en el programa.

En consecuencia puede haber razones endógenas y exógenas que expliquen que un programa no logre sus objetivos en la forma esperada, o que a pesar de lograrlos no contribuya a resolver el problema social que le dio origen.

Para identificarlas se realiza la evaluación de programas. Puede procederse con base en tres modelos diferentes de análisis:

- Evaluar únicamente algún elemento del proceso: sólo los procedimientos empleados en la gestión o evaluación en estudio, o bien, sólo los resultados alcanzados.
- Evaluar todo el proceso: diseño e implementación en sus diferentes componentes (insumos, procesos de transformación, resultados e impactos).
- Comenzar por la medición de impactos y, sólo si éstos no son satisfactorios, retroceder a analizar las brechas previas para encontrar las causas del fracaso y diseñar las correcciones posibles.

Debido a que una evaluación de impacto es difícil, costosa y sus resultados no suelen ser contundentes, a menudo se opta por el primer modelo de evaluación, en el que el impacto no se mide rigurosamente, pero se identifican las brechas que, dentro de la lógica del programa, se supone que deben estar disminuyendo la efectividad del mismo, y se procede a su corrección.

Si hubiera buenas razones para pensar que la medición de impacto se puede hacer con suficiente precisión, y motivos para creer que muchos de los programas aplicados son exitosos para la solución de problemas, se podría aplicar a todos el tercer modelo, procediendo a analizar todo el proceso sólo cuando se presenten resultados insatisfactorios, lo cual ahorraría bastante tiempo y dinero. En el desarrollo de este trabajo iremos viendo qué tan realista es pensar que podemos medir rigurosamente los impactos de un programa.

En términos generales, prácticamente todos los autores revisados esperan que la evaluación contribuya a la toma de decisiones y rendición de cuentas a la ciudadanía, coadyuvando a un empleo eficiente de los recursos y a una alta efectividad de su impacto. Algunos autores la consideran un componente crítico de la reforma del sector público (Ospina, 2001). Diversas perspectivas la asocian con el interés público, el mejoramiento de la sociedad y su contribución a la democracia (Scriven, 1967; Weiss, 1982). En la práctica, es necesario analizar en qué condiciones ayuda a realizar juicios y procesos de decisión más razonados y rigurosos (Thoenig, 2002).

### **Tipos de evaluación**

Existen muchas formas de clasificar a las evaluaciones (y metaevaluaciones), por lo que revisaremos sólo las que nos parecen de mayor utilidad para el presente trabajo. Se aclara que se consideran exclusivamente las evaluaciones realizadas en forma sistemática, objeto de nuestro trabajo, que implican una rigurosa recopilación de datos, basada en un buen conocimiento de epistemología, reglas empíricas, ciencias informativas, proyectos de investigación, valoraciones, estadísticas, comunicación, etcétera, descartando las evaluaciones efectuadas en forma intuitiva, según la clasificación de Stufflebean y Shinkfield (1987).

- a) Ya hemos hecho referencia a una primera clasificación en evaluación *ex-ante* o *a priori* y evaluación *ex-post* o *a posteriori*, vinculadas respectivamente al “conocimiento en” y al “conocimiento de”, de Laswell. Adicionalmente, hay autores que se refieren al monitoreo o evaluación *ex-tempore* o *concomitante* con el proceso de implementación; se trata en este caso de la utilización de herramientas propias del control y seguimiento de lo planeado. La evaluación

*ex-ante* se refiere a la comparación de los efectos esperados de las distintas alternativas que pueden permitir la resolución de un problema y constituye una de las etapas de un proceso decisorio que trabaja con base en datos estimados. En el presente trabajo, nos centraremos en la evaluación *ex-post*, de aquellas decisiones y acciones ya ejecutadas, que componen la gestión de los programas públicos seleccionados, por lo que sus efectos constituyen datos históricos. A pesar de que este tipo de evaluación identifique, mida, compare y califique los resultados o impactos obtenidos, eso no significa que se la deba considerar como la evaluación final propia de una actividad acabada, sino que puede constituirse en un proceso de seguimiento continuo de reflexión y análisis sobre los avances y limitaciones que se generan en el desarrollo de una actividad, lo que constituye una oportunidad para adoptar medidas correctivas oportunas y permitir un proceso de aprendizaje y crecimiento de las personas involucradas.

- b) La evaluación puede ser de carácter *externo*, en el caso en que es desarrollada por personas u organizaciones ajenas a quienes desarrollan la actividad evaluada. Puede tratarse de una evaluación *intergubernamental* (Rieper y Toulemonde, 1996), basada en modelos del tipo *top-down*, o *independiente*, desarrollada por una organización totalmente ajena al gobierno (centros de investigación, organizaciones no gubernamentales, periodistas, partidos políticos, sindicatos, etcétera). En el caso opuesto al anterior, se habla de una evaluación *interna*, como la que puede desarrollar el propio organismo responsable de un programa público. Si quienes hacen la evaluación son los mismos funcionarios que aplican el programa, o la población participante en el mismo, se trata de una *autoevaluación*.

Una evaluación interna garantiza un mayor conocimiento de la actividad revisada y disposición de información, mientras que una evaluación externa asegura mayor objetividad. Sin embargo, conviene resaltar que la evaluación no constituye un proceso libre de intereses, sino que por el contrario, existen personas y grupos preocupados porque los resultados que se obtengan demuestren el éxito de la gestión, mientras los opositores se obstinan en lograr lo contrario. Por esa razón, pueden existir incluso presiones sobre los evaluadores que provocan sesgos en los resultados, disminuyendo la “objetividad” que sería deseable en dicho proceso.

- c) Los destinatarios de la evaluación, pueden ser muy diversos. En la evaluación de programas públicos la mayoría de los destinatarios pertenecen a la

Administración Pública, que las realiza para retroalimentar su proceso decisorio; al Congreso, para servir de contrapeso del Ejecutivo; y a la ciudadanía, para fomentar el debate público y su participación. Algunos de ellos pueden involucrarse en la realización de la evaluación, que adquiere entonces, el carácter de *participativa*, misma que, cuando incluye a los ciudadanos interesados en el programa, contribuye a afianzar los procesos de democratización (Blanet y otros, 2000).

Esta evaluación, que Guba y Caracelli (1989) denominan “evaluación de cuarta generación” se organiza en torno a las peticiones, demandas, asuntos y cuestiones planteadas por los propios implicados, mismas que constituyen la base para determinar la información necesaria. Su metodología enfatiza el desarrollo de procesos de aprendizaje de los participantes (Isaac, 1999 y DeBono, 1999).<sup>2</sup>

Recientemente, Llerenas (2007) menciona que las deficiencias de un programa se detectan mejor y más rápido con la participación de los propios beneficiarios. En el mismo sentido, Guimaraes (2008), argumenta que los procesos de reforma y modernización del Estado, cuyas decisiones económicas se basan en una racionalidad técnico-instrumental, tienden a considerar imposible la incorporación de la participación; mientras él sostiene que es necesaria para la construcción de ciudadanía y la defensa de principios de solidaridad, equidad y democracia, por lo que analiza los requisitos que permitirían su concreción efectiva: capacidad de organización, derechos elementales de ciudadanía, estructura de oportunidades de participación, flexibilidad de la estructura institucional, intervención real en la toma de decisiones, entre otros (Cardozo, 2009).

d) Otra clasificación posible apareció insinuada al tratar el tema del momento en que se evalúa. Se trata de la que identifica los dos tipos siguientes: a) *sumatoria*,

<sup>2</sup> La participación social tiene un fuerte arraigo en los países europeos, especialmente los nórdicos, pero también se ha desarrollado en contextos muy diferentes. En Japón se le conoce a partir del concepto “ba”, que describe un lugar de aprendizaje, de carácter físico, virtual o mental, con el objeto de relacionar individuos u organizaciones, para facilitar el aprendizaje y desarrollo creativo, mediante procesos de socialización, externalización, combinación e internalización (Nonaka y Nishiguchi, 2001). Se trata de una forma de colaboración que, en Occidente, podría considerarse como formación de “capital humano” y “capital social”.

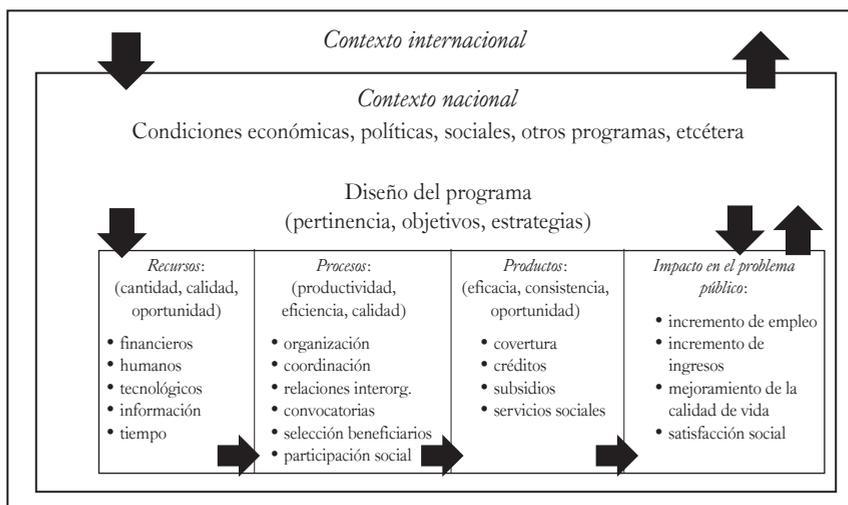
*de balance, recapitulativa o conclusiva*, y *b) formativa, endoformativa o correctora* (Scriven, 1967). La primera evalúa los resultados finales de una actividad, plan, política o programa aplicado, en comparación con los objetivos planeados o los resultados obtenidos en empresas similares. Constituye generalmente un instrumento de gestión, dirigido a personas ajenas al desarrollo de la actividad, que sirve para decidir si la misma debe proseguir, con o sin ajustes. La segunda realiza un seguimiento continuo de la actividad desarrollada para detectar a tiempo si se están presentando dificultades y tratar de resolverlas. Cuando se realiza en forma participativa, funciona como un instrumento de aprendizaje colectivo que permite a los mismos protagonistas de la actividad modificar conductas y mejorar acciones, contribuyendo así a los procesos de participación y democratización; en caso contrario, se limita al aprendizaje organizacional del gobierno (Leeuw y otros, 2000).

- e) El primer tipo de evaluación identificado por Scriven, enfatiza la cuantificación de los resultados hallados, mientras que el segundo hace su análisis en forma fundamentalmente cualitativa, se orienta hacia un conocimiento comprensivo de lo que está sucediendo. Esto nos genera una nueva clasificación, que nos marca diferencias entre una evaluación *cuantitativa* y *cualitativa*. Según Hofstède (1980), la evaluación cuantitativa sólo es posible cuando los objetivos perseguidos son claros y los resultados medibles y da lugar a controles de rutina, por expertos, por medio de pruebas e intuitivos. Cuando los objetivos son claros pero los resultados no son medibles, corresponde un control “a criterio” y, en los casos en que incluso la definición de objetivos es ambigua, el control sugerido es de tipo político. Estos últimos dos casos implican la realización de evaluaciones cualitativas, basadas en criterios personales y, por lo tanto, subjetivos.
- f) También suele diferenciarse entre evaluación *sectorial*, enfocada a las instituciones, planes o programas dirigidos a un área sustantiva de política, y evaluación *global*, que involucra diferentes sectores relacionados con una política.
- g) Stufflebean y Shinkfield (1987) hacen una diferenciación importante entre *pseudoevaluación*, *cuasievaluación* y *evaluación verdadera*. La primera intenta presentar conclusiones erróneas como fruto de un “evaluación”. Incluye las evaluaciones políticamente orientadas (partidistas, por ejemplo) o “investigaciones encubiertas”, que buscan un resultado positivo o negativo, independientemente de cualquier valoración objetiva. Lo mismo sucede con los estudios basados en las relaciones públicas, cuya intención es ayudar al cliente

a lograr una imagen positiva de algo. La siguiente, da respuesta a algunos asuntos de interés, en lugar de determinar el valor y el mérito de algo, objeto de la evaluación verdadera. Incluye los estudios basados en objetivos (Tyler) y en la experimentación que busca relaciones causales pero no emite juicios de valor. Las verdaderas evaluaciones se componen de estudios orientados a la toma de decisiones (Cronbach), los centrados en el cliente para valorar sus contribuciones, los estudios que miden los méritos relativos de diversas políticas y programas (Rice) y los que ayudan al consumidor a elegir los bienes y servicios más convenientes.

h) Finalmente, denominamos *evaluación integral* a la que cubre, utilizando el enfoque de sistemas, los insumos, el proceso de implementación, los resultados y los impactos, pero también el diseño y todo este conjunto inmerso en los contextos nacional e internacional (Diagrama 3). Dicha integralidad se complementa con el estudio pluridisciplinario de los elementos anteriores.

DIAGRAMA 3  
*Evaluación integral de un problema*



Fuente: elaboración propia.

Pero también existe un conjunto de trabajos (Nioche, 1982; Stufflebean y Shinkfield, 1987; Monnier, 1991 y 1995; Mény y Thoenig, 1992) que aportan la identificación de diversos tipos de *evaluaciones parciales*:

- de *objetivos*: analiza si éstos han sido definidos con precisión.
- de *pertinencia*: vincula los objetivos del programa con los problemas que los originan.
- de *coherencia*: relaciona los objetivos con los medios para alcanzarlos.
- de *contexto*: revisa las condiciones en que se ha producido la aplicación de la política o programa así como los procesos de interacción del programa y su entorno.
- de *insumos* o de *medios*: analiza la cantidad y calidad de los recursos empleados y verifica si los mismos han estado disponibles en el lugar y tiempo preestablecidos.
- de *proceso*: revisa los mecanismos que permitieron la transformación de insumos en resultados.
- de *realizaciones, resultados o productos*: intenta apreciar los productos o servicios generados por la actividad evaluada.
- de *eficacia*: verifica si los resultados han permitido lograr los objetivos.
- de *metas*: analiza lo logrado en comparación con lo planeado.
- de *eficiencia*: relaciona los resultados con los esfuerzos realizados para lograrlos (costo-beneficio, costo-eficacia).
- de *efectos*: mide las consecuencias provocadas por los resultados alcanzados en el entorno socio-económico, sin juzgar los objetivos del decisor, incluyendo las externalidades provocadas, en tanto efectos no previstos, positivos o negativos; busca apreciar las modificaciones de la situación o de los comportamientos de los individuos, empresas o comunidades afectadas por la política. Los efectos inesperados se producen debido a cambios del entorno, errores en la estructuración causal de la acción o grandes consecuencias colectivas provocadas por pequeños efectos (“efecto mariposa”), y demuestran la ausencia de linealidad causa-efecto de la acción colectiva.
- de *impacto*: se refiere al impacto global, provocado por la combinación de efectos atribuibles al programa, sobre la problemática que justifica la existencia de la actividad.

- de *satisfacción*: determina el grado en que la población considera satisfecha su necesidad o resuelto su problema.

En materia de medición de impacto, si bien el aparato administrativo público está fuertemente diferenciado y cuenta con una débil capacidad de integración, es necesario señalar que la transversalidad de las políticas obliga a realizar evaluaciones que no respetan los cortes institucionales, requiriendo un mínimo de acuerdo y cooperación entre los actores y los niveles de acción y responsabilidad (Duran, 1999).

### **Actitudes y enfoques**

- a) Una actitud que comúnmente se espera de un evaluador es que realice su trabajo en forma tan *objetiva* como sea posible. Este es un aspecto muy importante dentro de ciertos enfoques metodológicos (positivismo, conductismo), muy cuestionable y poco realista para otros (relativismo) y una falsa dicotomía para Pierre Bourdieu. En el caso concreto de la evaluación, aún suponiendo que pueda mantenerse un alto nivel de objetividad en la consecución de los cuatro primeros pasos que la conforman (identificación, medición, comparación de efectos y explicación), el quinto implica un juicio de valor que, por mucho que se fundamente en criterios preestablecidos, no dejará de tener elementos de carácter subjetivo. Lo mismo sucede con el sexto paso, aunque probablemente en menor grado. Adicionalmente, pueden existir durante el proceso de evaluación diversos tipos de compromisos que disminuyan aún más dicha objetividad: afectivos, de asociación, factuales, fiscales, filosóficos y políticos, que deben tratar de evitarse.

En ese sentido, Duran expresa: “Si hasta aquí se ha tratado de consagrar la autonomía de las esferas de los valores y los hechos, la evaluación viene a recordar su interdependencia [...] La política sigue siendo asunto de juicio y compromiso” (1999:188, traducción libre).

Esta combinación objetiva-subjetiva de la evaluación hace que su credibilidad dependa no sólo de la *expertise* del evaluador, sino también de la confianza que éste genera con base en sus características morales, que le permitan un manejo honesto de la información y una emisión no

comprometida de su juicio valorativo; mismo que resulta más garantizado cuando la evaluación queda a cargo de una instancia colectiva y plural, como veremos más adelante. La deontología, basada en los principios de autonomía financiera e institucional del evaluador para garantizar su objetividad, y, de replicabilidad posible de los resultados por otro evaluador que utilice la misma metodología, aparece como un tema central de la evaluación desde el trabajo de Deleau y otros (1986), que la coloca al mismo nivel que los problemas vinculados con su método y forma de organización.

García (2009) plantea seis razones para destacar el carácter político, no objetivo, de toda evaluación: 1) las políticas y los programas públicos sometidos a la evaluación son aprobados y financiados dentro de un proceso político, 2) la evaluación reproduce los rasgos del contexto político en que surge y sus estructuras, actores y procedimientos influyen fuertemente en su realización, 3) los gobiernos ven en ella una oportunidad de control y un elemento de legitimación, razón por la que la impulsan y financian, 4) en la evaluación misma se expresan intereses políticos contrapuestos, que buscan aumentar sus cuotas de poder e influencia, 5) el diseño de la evaluación y sus criterios son de naturaleza política en la medida en que implican tomar posición sobre la participación de grupos en la misma o la apropiación de la información generada, y 6) se realizan para influir en el proceso de toma de decisiones política, para ganar legitimidad y fundamentar argumentos, entre otros.

De esta forma, García señala que varios estudios han servido para mostrar el mal estado en que se encuentra la evaluación financiada con presupuesto público en términos de credibilidad ante la injerencia creciente del gobierno en la elaboración de la propuesta y la imposición de restricciones al uso y difusión de resultados, establecidas en el contrato.

b) Mény y Thoenig (1995) identifican cuatro posibles actitudes en un evaluador: *descriptiva, clínica, normativa y experimental*. La actitud descriptiva es la que adopta un evaluador que pretende hacer un listado exhaustivo y neutral de los efectos de una intervención, sin emitir ninguna valoración. Más que un evaluador, se trataría de un informante. En cambio, un evaluador con actitud clínica, complementará la descripción anterior con una explicación relativa al nivel de logro de los objetivos perseguidos por la gestión, basada en el análisis de las condiciones de la ejecución de la misma. A diferencia del anterior, un

evaluador con actitud normativa analiza los resultados encontrados a la luz de sus propios valores, que pueden diferir de los que el programa pretendió alcanzar, entra al debate político y normativo, al tener que tomar partido por alguno de los valores, datos y plazos que existen en la sociedad. Finalmente, la actitud experimental es la más ambiciosa desde el punto de vista científico ya que pretende establecer rigurosas relaciones de causalidad entre la gestión en estudio y los efectos que se le imputan.

Mény y Thoenig se refieren también a la existencia de evaluaciones *como actividad política*, donde la información es intencionalmente descartada o tergiversada a los efectos de que se ajuste a los prejuicios culturales, intereses y marcos cognitivos de cada persona; por lo que, en estos casos, la información no constituye la variable esencial para la formulación de los juicios de valor.

- c) En materia de enfoques, la evaluación sufre la influencia del debate epistemológico ente corrientes positivistas y constructivistas. Los *enfoques positivistas* se caracterizan por poner distancia respecto del objeto, preocuparse de la validez interna y externa de su diseño metodológico, la confiabilidad, el poder de generalización, la objetividad, el énfasis en la cuantificación y la utilización privilegiada de la experimentación.

Los *construccionistas* (R. Mayntz y la Escuela de Colonia, Duran y Thoenig) creen en la existencia de múltiples realidades (en lugar de visiones objetivas), se interesan en fenómenos que articulan lo micro y lo macro, lo formal y lo informal, la organización en red, rechazan la separación tajante de las etapas del proceso de políticas públicas y la realización de estudios de caso sectoriales, y enfatizan la investigación cualitativa, a veces apoyada en la estadística.

Monnier (1989 y 1995) considera que es posible clasificar los enfoques de evaluación en *tradicionales* (jurídico y técnico-económico), basados en el positivismo, y *nuevos métodos* (endoformativos, pluralistas y efectuados por los administrados), fundamentados en el constructivismo.

El modelo jurídico comenzó aplicando el control jerárquico de las prácticas administrativas, al que luego se agregaron los aspectos técnicos y financieros, manteniendo la intención de comprobar, por parte de alguna instancia de la elite administrativa, si la acción realizada se ajustó a los medios y procedimientos definidos por las reglas preestablecidas. No intenta medir la eficacia de la acción, mucho menos su impacto social.

El segundo modelo equipara la gestión pública y la privada. En materia de evaluación, supone la existencia de objetivos explícitos y contextos estables, enfatiza lo cuantificable y excluye todo elemento ajeno a lo técnico-económico. Se desentiende de la dinámica real de los procesos, por lo que genera efectos perversos en entornos de alta incertidumbre. Mejía (2003) le identifica dos limitaciones principales: la irrelevancia de una evaluación que no responde a las necesidades de los actores-clave (gerenciales o sociales) y su limitada institucionalidad, debido a que goza de un débil apoyo gubernamental y no gubernamental.

En cambio, Monnier propone enfocar la evaluación hacia el sistema de actores sociales involucrados, con una actitud *pluralista*, cuyo antecedente lo constituyen las evaluaciones participativas, denominadas de Cuarta Generación por Guba y Caracelli (1989). Dicha actitud se fundamenta en el reconocimiento de la diversidad de sistemas de valores que coexisten en el seno de la sociedad y se construye a partir de los conflictos sociales emergentes; lo que lleva al evaluador a actuar como conciliador entre las partes en desacuerdo, en búsqueda de una solución satisfactoria, no óptima.

De manera cercana, Stake (2010) opina que la evaluación depende de quien la está realizando. Reacciona ante los enfoques evaluativos, supuestamente objetivos y cuantitativos, basados en criterios y estándares preestablecidos para apostar por una evaluación comprensiva o interpretativa que utiliza métodos cualitativos que recuperan la subjetividad de los actores, la particularidad de las situaciones y sus contextos (Stake, 2006).

Monnier (1991) plantea, además, que la utilidad social de una evaluación depende, no sólo de la metodología empleada, sino de dimensiones políticas como el acceso a la información, la existencia de condiciones que favorezcan la utilización de sus conclusiones y la credibilidad de la misma para los distintos actores, determinada por la legitimidad del proceso seguido, ya que ninguna evaluación puede considerarse neutral.

Desde el enfoque pluralista, Monnier afirma que “un evaluador experimentado reúne aptitudes, a nuestro juicio, para llevar a cabo una evaluación en un sector que no le resulte familiar. Su desconocimiento le permite salvaguardar la disponibilidad mental indispensable” (Monnier, 1995:271); en cambio, Mény y Thoenig (1992) consideran que el evaluador debe ser un experto en el tema objeto de su evaluación.

- d) En el mismo sentido, Duran (1999) señala que la mayoría de quienes solicitan una evaluación pretenden conocer todos los efectos de una política o programa y poner en evidencia los esquemas de causalidad que los producen. En función de ello el rigor metodológico se ha confundido con un enfoque *tecnicista*, descuidado la construcción del objeto del estudio y la elaboración de un consecuente protocolo de investigación. Se olvida que las ciencias se limitan a la aplicación de métodos y técnicas y que, aunque la verdad sea única, existen múltiples puntos de vista en un trabajo interdisciplinario y participativo, por lo que los primeros no garantizan totalmente la pertinencia científica de las conclusiones. Debido a que la ciencia no cuenta más que con procedimientos de verificación limitada, es que proliferan más las evaluaciones de proceso que de resultados.
- e) Hace ya más de 20 años, Nioche planteaba la conveniencia de avanzar del enfoque de *evaluación-sanción* al *análisis-aprendizaje*. Duran y Monnier (1992) también se han declarado a favor del aprendizaje a partir de la experiencia de enfrentar problemas mediante la aplicación de soluciones nuevas, en lugar de intentar exhaustivos ejercicios de programación de objetivos, metas, estrategias, etcétera. Se requiere de una cierta indeterminación, por tanto de libertad, para que se pueda actuar en forma estratégica (Duran, 1999).

Engel y Carlsson (2002), por su parte, plantean que existen tres distintas perspectivas para impulsar el aprendizaje desde las evaluaciones, a diferentes niveles de la complejidad y participación social: 1) el enfoque tradicional de tomadores de decisiones con participación simbólica para legitimar lo realizado, 2) el aprendizaje organizacional de funcionarios, financiadores y demás actores relevantes, y 3) el impulso al aprendizaje de la sociedad, considerado por el BM como un bien público global. En estos casos, el evaluador debe recoger y juzgar evidencias, pero también actuar como facilitador de un proceso de aprendizaje, por lo que, implica un cambio de rol y de las características requeridas para cumplirlo (interdisciplinarietà, habilidades de negociación, didácticas, etcétera).

En el contexto mexicano prevaleció durante el siglo pasado un enfoque punitivo, que poco a poco se va sustituyendo por otro centrado en el aprendizaje de funcionarios, legisladores y ciudadanos en general, Ochoa (2007, citada en Cardozo, 2009) define a la evaluación en ese sentido, como un proceso de aprendizaje continuo que permite conocer los resultados de la

gestión, que identifica aciertos y errores para que, con base en esa información, se realicen los cambios pertinentes en la gestión. Insiste también en una justificación de la evaluación que vaya más allá de las razones técnicas: evitar el uso de programas con fines políticos y electorales, así como la corrupción y el desvío de fondos.

Obviamente que la participación y el aprendizaje social no están exentos de riesgos, como algunos casos en que han provocado resultados no deseados, lo que lleva a la existencia de autores que se oponen terminantemente a ella (Cooke y Kothary, 2001).

f) Sulbrandt (1993) clasifica los enfoques de la evaluación en: 1) *tradicionales* y 2) *evaluación global*. Considera que, salvo excepciones, la evaluación tradicional no ha satisfecho las expectativas en el caso de los programas sociales, de características muy diferentes a los dirigidos a la inversión económica, y suele ocultar aspectos desfavorables al programa, a la gerencia o a la institución encargada de implementarlo; en consecuencia, no ha conducido a provocar ningún cambio en el manejo de los programas ni en el aprendizaje de los ciudadanos. Además, sus resultados sólo están disponibles para quienes encargan la evaluación, aparte de los múltiples problemas técnicos que se le critican (imposibilidad de atribuir causalmente los resultados al programa, falta de oportunidad, poca relevancia y utilidad, perspectiva de “arriba” a “abajo”, etcétera).

En contraposición a la evaluación antes analizada, Sulbrandt propone una evaluación global, que buscaría constituirse en una instancia de aprendizaje que permitiera rediseñar y mejorar los programas, a partir de un trabajo cooperativo, aprendiendo de los errores, omisiones y dificultades encontradas en su gestión.

La propuesta consiste en llenar los huecos dejados por el enfoque tradicional, particularmente: 1) contar con una elaboración participativa de criterios de los diferentes actores sociales involucrados, 2) tomar en cuenta lo sucedido durante la etapa de implementación, 3) identificar el grupo-objetivo, funcionarios encargados y demás actores con sus respectivos intereses en el programa, frente a su total omisión, y 4) comprender la lógica interna del programa y sus bases teóricas (cadenas causales).

El nuevo enfoque utiliza datos objetivos y subjetivos, cuantitativos y cualitativos, mezclando una variedad de métodos de recolección de datos,

con objeto de aumentar la comprensión del programa, así como la legitimidad de sus resultados.

Para lograr lo anterior se necesita prestar mucha atención a la capacitación de evaluadores competentes e institucionalizar a la evaluación dentro de la administración pública.

Hace ya una década, el BM (2001) incluyó el trabajo denominado “Voices of the poor” en uno de sus reportes, que recupera la participación, en discusiones en pequeños grupos y entrevistas grupales e individuales, de 60 000 personas pobres en 60 países. También el Banco Interamericano de Desarrollo (BID, 2000), la OCDE (2001) y otras agencias de desarrollo impulsan el aprendizaje participativo a partir del análisis de los éxitos y fracasos de acciones y programas.

En México, la Secretaría de Desarrollo Social (Sedesol) también aplicó en 2003 una encuesta denominada “Lo que dicen los pobres”, dirigida a captar las percepciones y expectativas de la población que vive en diferentes niveles de pobreza sobre bienestar, vida digna, satisfactores mínimos, entre otros (Székely, 2003).

Como puede verse, los temas de la participación y el aprendizaje social aparecen en la caracterización de diversas actitudes que se oponen a los enfoques tradicionales, constituidos por evaluaciones gubernamentales con el objetivo de retroalimentar la toma de decisiones, sin tomar en cuenta el punto de vista de los actores sociales ni informarles de los resultados hallados. Dan lugar a tres posibles niveles de participación de la sociedad civil en la evaluación:

- Evaluaciones similares a las tradicionales, pero que incluyen la aplicación de encuestas con preguntas preestablecidas para conocer algunos datos y opiniones de los actores sociales,
- Evaluaciones realizadas en la forma anterior, que agregan el empleo de procedimientos de rendición de cuentas a la ciudadanía de los resultados encontrados (Nioche), y
- Evaluaciones realizadas con la participación de los actores sociales (Monnier).

g) Actualmente, en Europa (EES, 2002), se habla de tres enfoques alternativos (a pesar de que no siempre se ven tan independientes) dentro de la evaluación

actual: *aprendizaje, teorías y pruebas*. Se trata de una discusión que no está presente en el contexto latinoamericano. El primero se refiere a los procesos ya comentados de aprendizaje a partir de las evaluaciones pasadas, a nivel gubernamental y de la sociedad civil. El segundo recupera las evaluaciones basadas en la teoría (*theory-based approach*), más allá de la mera aplicación de buenas técnicas y métodos, que no siempre ofrecen resultados claros (Huey Chen, Carol Weiss, etcétera); esta corriente se ha fortalecido con el reciente surgimiento de escuelas “realistas”. Finalmente, el tercero retoma viejas corrientes que buscaron darle un enfoque científico al proceso político, trabajando con base en evidencias logradas en pruebas experimentales rigurosas y normas estrictas (Cochrane, Campbell, etcétera).

Muy brevemente diremos también que hay un enfoque *territorial*, local, de la evaluación de la acción pública. Se trata de evaluar una mezcla de políticas y/o programas aplicados a un mismo territorio. Su ventaja sería que evita el problema de aislar los efectos de cada política, manteniendo sólo las fronteras de los espacios geográficos en que se aplican. La complicación es que todos los programas pueden adjudicarse los resultados positivos y desviar hacia los otros los negativos, o sea no hay una clara responsabilización de los impactos.

Una síntesis de los principales cambios teórico-conceptuales que se ha experimentado en los países ubicados en la vanguardia evaluativa es ofrecida por House (1993, citado en Dopico, s/f): sustitución de nociones monolíticas por otras pluralistas, abandono de la idea de una evaluación libre de valores, mayor énfasis en la concepción “iluminativa” que en la instrumental e insistencia en el carácter político de aquélla (Cardozo, 2009).

Finalmente, a cada actitud o enfoque corresponden diferentes tipos y métodos de evaluación a emplear, que se revisarán en el siguiente capítulo.

### **Contexto político-organizativo de la evaluación**

Si bien el problema del método es un asunto clave en materia de evaluación, hay otro aspecto muy importante que enmarca la forma en que se desarrollan sus trabajos: las dimensiones organizacionales y contextuales. Se trata de comprender

que las administraciones públicas, en cuyo marco se realizan los procesos de gestión y evaluación, se componen de redes de organizaciones en que los factores políticos, culturales, de estructuración y de funcionamiento constituyen factores relevantes que afectan el desempeño evaluativo. El estudio de dichas condiciones, uno de los tres resaltados<sup>3</sup> en Francia por el Informe Deleau en 1986, constituye el enfoque aportado por la sociología de las organizaciones y la sociología política.

Duran y Monnier (1992) plantean que, efectivamente, la complejización de la acción pública y las dificultades del aparato estatal para justificarse son dos de las principales razones que explican el desarrollo de la evaluación, junto con el avance del análisis de políticas públicas. Se refieren a una crisis del sistema de representación en que la política ha entrado en la “era de la sospecha”, signo del deterioro de la democracia representativa como expresión de preferencias, por lo que los gobiernos tratan de resolver problemas concretos al mismo tiempo que de recuperar el mayor consenso posible.

Pero, legitimidad y eficiencia son difíciles de conciliar. Antes el Estado francés y el interés general concentraban místicamente toda legitimidad, actualmente ésta resulta aleatoria y contingente, y debe ganarse mediante la acción eficaz, de la visibilidad de las consecuencias de las elecciones realizadas. Así la evaluación produce al mismo tiempo conocimiento y justificación (Duran, 1999).

Chanut (2002), quien fuera miembro del Consejo Nacional de Evaluación en Francia, resalta la importancia de la dimensión organizacional en los procesos de evaluación y considera que el énfasis puesto en los temas de metodología la ha mantenido oculta y también ha impedido la capacidad de innovación en la materia. Su trabajo plantea:

1. Las organizaciones se mueven en un mundo de restricciones que afectan a la evaluación: diferencias entre la lógica del conocimiento social y la burocrática dirigida a la acción, incompatibilidad de los tiempos de ambas, falta de acceso a la información, ambigüedad por desaparición de fronteras público-privado, participación de nuevos actores, *expertise* de la oferta de evaluadores, procedimientos impuestos, presupuestos asignados, etcétera.

<sup>3</sup> Los otros dos temas eran metodología y deontología (reglas del arte).

2. La evaluación no ha conseguido aclarar los problemas y confusiones que disminuyen la eficacia organizacional: pérdida de legitimidad, obsolescencia de los dispositivos e institucionalización de prácticas estereotipadas. La credibilidad de las evaluaciones se funda en la autoridad de los evaluadores y se legitima en el marco de estructuras muy dependientes de la turbulencia política, por lo que la obsolescencia acelerada de los programas provoca una débil capitalización de experiencias. Las formas organizacionales determinan el tipo de evaluación realizada y la naturaleza de los resultados: multiplicación de sus objetivos en función del pluralismo; tendencia a la medición de resultados útiles a la gestión administrativa pero descuidando la evaluación de impacto; énfasis en la implementación de políticas y sus disfuncionamientos, que descuidan la concepción misma de la política; muy pocas experiencias de metaevaluación para la capitalización y apropiación de sus resultados; etcétera.
3. Una mayor atención a los problemas organizacionales aumenta la credibilidad y la eficacia de la evaluación. Propone organizar la evaluación de manera heurística, en lugar de procedimientos y métodos excesivamente formalizados; y pragmática, pasando de constituir una amenaza para los evaluados a transformarse en una fuente de información y aprendizaje útil para la acción, mediante una verdadera restitución y discusión de los resultados por los actores, en lugar de la mera entrega burocrática. Considera que de su adaptación depende la legitimidad y sobrevivencia de la evaluación.

En el mismo sentido Hespel, integrante del Comisariado General del Plan, reconoce que “la consolidación de la evaluación dependerá estrechamente de los medios y competencias concretamente aplicados y de la manera como los actores se apropien de sus enseñanzas” (Hespel, 2002:51, traducción libre).

Lo anterior no es sencillo puesto que los decisores no actúan con una lógica de consecuencias exógenas, sino de acuerdo con una cultura compartida, compuesta de reglas internas y modos de razonamiento implícitos. Esto provoca que los errores de cálculo y las consecuencias no deseadas sean los elementos constitutivos de la acción pública, no aberraciones marginales. Por lo que se propone redefinir la evaluación como un proceso de aprendizaje institucional, de transformación de lo que conviene hacer, de las reglas y las identidades. Para institucionalizar realmente esa evaluación, se señala que es necesario algo muy

diferente a establecer procedimientos administrativos: crear las condiciones que se enfocan a cambiar la teoría en actos de los decisores, a modificar los modelos cognitivos y normativos que estructuran su visión del mundo, su representación de los problemas y su apreciación de los instrumentos movilizados para actuar (Thoenig, 2002).

### **Comentario final**

Más allá de la gran diversidad de definiciones, tipos y enfoques de evaluación revisados es posible identificar dos tendencias fundamentales:

- Una, basada en un enfoque positivista, preocupada por la objetividad del trabajo, que se compone de definiciones de tipo instrumental, interesada fundamentalmente en la evaluación de tipo sumatoria y desarrollada desde un enfoque no participativo.
- Otra, fundamentada en el constructivismo, motivada por la credibilidad y legitimidad del proceso y sus resultados, que aporta definiciones de corte sociológico, y prioriza las aplicaciones de evaluaciones formativas con enfoque participativo e interés en el aprendizaje social.

Cada una se acompaña también de énfasis diferentes en cuanto a la preocupación y selección de los métodos y técnicas de evaluación, que serán objeto de exposición en el próximo capítulo.

## CAPÍTULO III

### Métodos y técnicas de evaluación

Es el momento de referirnos a los métodos y técnicas empleados en la evaluación *a posteriori*, así como a sus respectivos criterios de decisión. Sin embargo, no se realizará una presentación exhaustiva de todos los métodos, lo que excedería las posibilidades de este trabajo, sino que se mencionarán las principales características, aportes y limitaciones, de cada uno de ellos y se ofrecerán referencias bibliográficas donde los interesados podrán avanzar en su conocimiento.

#### **Debate sobre el método**

El tema del método de evaluación es un reflejo de toda la discusión epistemológica generada en torno a las posibilidades y formas de acceder al conocimiento que, en nuestro caso, opone fundamentalmente enfoques positivistas y construccionistas (relativistas), pasando por una perspectiva contingente, que utiliza a cualquiera de los anteriores o incluso una mezcla de ambos, dependiendo del contexto de la evaluación.

Beltrán (1991) plantea que, tanto teórica como prácticamente, no existe unanimidad en lo que se entiende por método científico, en tanto método único, detallado, riguroso y universalmente aceptado; sólo una aceptación general de sus principios: contrastación como evidencia empírica, juego entre teoría y realidad, exclusión de la manipulación y el engaño, recurso a la comunidad científica como árbitro y reconocedor de la verdad, y renuncia a la justificación

absoluta de la verdad encontrada. Agrega la inconveniencia de que las ciencias sociales tomen como modelo a las ciencias naturales debido a las peculiaridades de su objeto.

Al hablar de métodos nos referimos a procedimientos de investigación que se desprenden de una postura epistemológica y un marco teórico, en que el evaluador, basado en esta perspectiva y a partir de un conjunto de principios y normas, realiza una serie de operaciones que le permiten seleccionar y coordinar ciertas técnicas para alcanzar sus objetivos. Estas últimas constituyen instrumentos concretos de recolección, análisis, tratamiento, síntesis e interpretación de datos,<sup>1</sup> que aparecen mejor desarrolladas que los primeros, los que requieren de una mayor discusión crítica, poco recogida en la literatura sobre evaluación. Esto sucede incluso entre constructivistas (Cook y Campbell, Cronbach), que ponen el acento en los efectos de contexto, resaltando la pluridimensionalidad de los problemas objeto de análisis y la diversidad de enfoques de los actores implicados en la evaluación.

Desde una actitud pluralista, se insiste en que la metodología no se reduce a la selección de instrumentos, “implica una reflexión sobre el conjunto de procesos que permiten, en una sucesión de interacciones y decisiones más o menos conflictivas, la producción de la evaluación por los actores sociales, institucionales, políticos, administrativos, etcétera” (Lascoumes y Setbon, 1996:6; traducción libre), por lo que proponen pasar del método como herramienta al método considerado como proceso de construcción, que puede ser analizado sociológicamente.

También resulta preocupante la tendencia a aplicar las mismas recetas metodológicas en diferentes contextos, obteniendo conclusiones a partir del mero tratamiento de los datos, sin tomar en cuenta una teoría explicativa que permita su interpretación (Thoenig, 2002). Identificar la teoría del cambio social implícita en toda política pública es esencial para el evaluador, para comprender la relación de causalidad que se pretende accionar (Mény y Thoenig, 1996).

La evaluación de políticas públicas consiste en poner a prueba, como si se tratara de una hipótesis, explícita o implícita, si una política o programa de acción

<sup>1</sup> Adaptado de las notas suministradas por el profesor Gilberto Jiménez, durante sus conferencias sobre epistemología de las ciencias sociales, Facultad de Ciencias Políticas y Sociales, Universidad Nacional Autónoma de México, 20 y 23 de enero de 2003.

determinados han sido o no capaces de provocar los efectos esperados. Esto implica asegurarse de que dichos efectos se han producido, y que la política ha sido su única causa, descartando impactos de otras posibles variables intervinientes.

Corresponde aclarar que los métodos y técnicas de evaluación que revisaremos son utilizados tanto en la evaluación de políticas como en la de programas públicos, en parte por la dificultad práctica de diferenciarlos en muchos casos concretos, y en parte, como señala Weiss (1998), porque han resultado bastante adaptables a cualquiera de ellos.

La forma en que la evaluación se lleva a cabo, depende de las múltiples combinaciones de respuestas a las preguntas formuladas en el capítulo anterior, razón por la que hay un gran número de posibilidades para responder a necesidades diferentes. En particular, la discusión sobre la utilización de métodos cuantitativos o cualitativos nace con los orígenes mismos de las ciencias sociales y las propuestas alternativas de la corriente positivista (Durkheim) e interpretacionista (Weber); sin embargo, se debe resaltar que ambos autores emplearon los dos tipos de métodos según lo requiriera su objeto de estudio, por lo que no debe confundirse el uso de un método con la adhesión a una escuela teórica.

Los debates entre comunidades que se identifican con ciertos métodos expresan aspectos más profundos, no siempre explícitos, porque indirectamente se vinculan con el compromiso con ciertos supuestos subyacentes en los paradigmas o con alguna concepción teórica (Tarrés, 2001:40).

A pesar de los esfuerzos en la última década por reconciliar estas posturas, la controversia todavía muestra la presencia de comunidades científicas empeñadas en una especie de dogmatismo metodológico.

A continuación se exponen brevemente los principales métodos que la evaluación de políticas públicas ha tomado de la investigación en ciencias sociales.

### **Experimentación y métodos cuantitativos**

Como se recordará, entre las posibles actitudes del evaluador se incluyó la denominada experimental, que se concreta en el método que lleva el mismo nombre. Dicho *método experimental* busca establecer relaciones de causalidad

entre el programa y su impacto y toma como referencia el paradigma metodológico desarrollado para las ciencias naturales, y aplicado desde hace 50 años por la epidemiología.<sup>2</sup> El mismo ha tratado de adaptarse al contexto de las ciencias sociales debido al alto nivel de rigurosidad que le permite detentar la exclusividad de verificar relaciones de causa-efecto estables entre ciertas acciones y sus supuestas consecuencias (verificación no aceptada como posible por corrientes epistemológicas como el relativismo). Se ha aplicado más a la evaluación de programas que a la vinculada con políticas globales.

Para poner en práctica un experimento social, se requiere construir un estado contrafactual que estime lo que habría ocurrido de no haberse realizado la intervención en cuestión (BM, 2004a:237). Para ello se necesita diseñar la investigación evaluativa desde antes de dar inicio a las acciones o programas en consideración. Dicho diseño, implica conformar al azar dos grupos entre los destinatarios de las actividades evaluadas, lo que garantiza estadísticamente su equivalencia. En estos dos grupos, supuestamente similares en sus características, se introduce una sola diferencia: la acción, plan o programa a evaluar que se aplica en uno (grupo-caso o tratamiento) y no en el otro (grupo-control o testigo). Lo anterior permite hacer mediciones iniciales y finales (e intermedias si se desea mayor control) del asunto de interés, en ambos grupos, e identificar como causa de las diferentes evoluciones seguidas por los mismos a la actividad desarrollada sólo en el grupo-caso.

El diseño experimental tiene la ventaja de permitir la observación de la diferencia entre las mediciones “antes-después” del programa ( $C - A$  en la Gráfica 1), comparada con la evolución sufrida por grupos de similares características iniciales (punto de partida aproximadamente coincidente en la gráfica) a los que el programa no se aplicó pero que, por otro tipo de razones ajenas al programa mismo, podrían también haber evolucionado, favorable o desfavorablemente ( $B - A$ ), lo que a su vez permite identificar el verdadero impacto provocado por las acciones evaluadas ( $C - B$  en la gráfica).

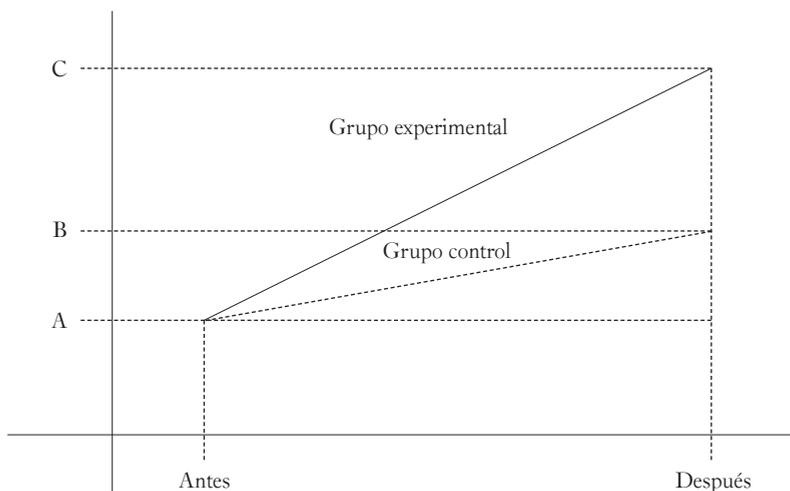
Este plan experimental debe cumplir dos criterios (Mény y Thoenig, 1992): una validez interna, lograda mediante la disminución de la posible influencia de

<sup>2</sup> Consultar Schwartz, “Les Techniques de l'évaluation en épidémiologie”, en Nioche y Poinard (1985) sobre el desarrollo de ensayos terapéuticos y de intervención, mediante la realización de una observación simple, una experimentación sistemática y una con elección al azar.

factores extraños en la evaluación registrada; y una validez externa, que permita la generalización de resultados a otras políticas o poblaciones supuestamente homogéneas.

La conformación de grupos equivalentes en ciencias sociales, se ha dificultado notoriamente tanto por problemas técnicos como éticos, razón por la cual se ha pensado en diseñar una metodología más flexible, menos rigurosa pero más realista. Surgen así los *cuasi-experimentos* en que se mantienen las características generales del diseño anterior pero se acepta que los grupos se constituyan en la forma más parecida posible (por pares, por ejemplo), renunciando a la equivalencia fundada en la selección aleatoria. Obviamente, aparece el riesgo de que variables extrañas, no controladas en el cuasi-experimento, afecten los resultados, induciendo errores. Por esto se recomienda hacer un esfuerzo por analizar y descartar esas posibles influencias antes de concluir la probable existencia de una relación de causalidad entre la acción y el supuesto efecto (Campbell y Stanley, 2005).

GRÁFICA 1  
*Medición de impacto del programa*



Fuente: elaboración propia.

A menudo, tampoco el método anterior es factible de ser aplicado, ya sea porque no contamos ni siquiera con grupo-testigo no-equivalente, por lo que no es posible comparar resultados; o porque carecemos de información previa al inicio de la acción a evaluar, por lo que no es posible analizar evoluciones “antes-después”. Se habla entonces de la aplicación de *métodos no experimentales* del tipo “antes-después sin grupo-control”, “después solamente con grupo-control” o “después solamente sin grupo-control”, en que cada uno resulta menos riguroso que el anterior y sus conclusiones menos precisas (Weiss, 1987).

Los análisis “antes-después” de un mismo grupo (corte longitudinal) o las comparaciones entre grupos (mediante cortes transversales), se realizan a partir de la generación de indicadores para las variables impactadas, siguiendo la lógica de la investigación aplicada (Cohen y Franco, 1993) o el Marco Lógico aplicado por los organismos internacionales.

Ante la imposibilidad de medir todos los aspectos deseados con una misma unidad de medida (tipo valor monetario), es necesario realizar un análisis segmentado de las distintas actividades y aspecto del logro. Para ello se utiliza un diseño analítico de *indicadores* parciales, mismo que se trata de complementar con un índice integrado, que ofrezca una comprensión sintética del logro, para lo cual se necesita tener claridad respecto de la jerarquía de objetivos que permita la adecuada ponderación de su valor (Abate, 2000).

Los indicadores se definen como la expresión de medida de un atributo (Sanín, 1999), medidas cuantitativas de aspectos del logro, a menudo expresadas como radios o ratios, que permiten contar con señales sobre su desempeño (OCDE, 1995), mediante su comparación con los referentes internos o externos. Constituyen instrumentos para comprobar si hay o no un acercamiento progresivo a las metas propuestas y permiten operativizar la medición de las variables, señalando sus grados y variaciones en la forma más objetiva y verificable posible.

Su diseño debe cumplir al menos tres requisitos:

- 1) permitir comparaciones sincrónicas, en el espacio, 2) permitir comparaciones diacrónicas, en el tiempo, y la elaboración de pronósticos, y 3) representar valores sobre los que existe un grado de consenso (o posibilidad de lograrlo) entre las partes implicadas o interesadas en la evaluación (Ander-Egg y Aguilar, 1998:64).

También se le exigen varias condiciones generales: independencia, verificabilidad, validez y accesibilidad. A éstas se agregan la posibilidad de ser medidos (cuantitativa o cualitativamente), contar con información disponible, ser determinantes, o sea, los que expresen mejor la naturaleza del objeto medido (Sanín, 1999) y contar con relevancia, falta de ambigüedad, pertinencia, objetividad, sensibilidad y precisión (Guinart i Solá, 2003).

Como puede verse, el conjunto de requisitos y condiciones es tan amplio que dificulta el adecuado diseño de los indicadores. Incluso, a los anteriores podríamos agregar dos restricciones opuestas que es necesario conciliar: 1) la integralidad o suficiencia, es decir, la necesidad de que el conjunto de indicadores diseñados dé una cobertura razonable a todas las variables identificadas como claves, y 2) su número limitado para no caer en una construcción que se vuelva inmanejable. Su combinación dependerá de la calidad y cantidad de datos disponibles, de los recursos asignados a la evaluación y del monitoreo previsto en el diseño del programa.

La construcción de indicadores requiere el conocimiento de los objetivos y actividades de la organización, política o programa evaluado, las fuentes de información disponibles, la periodicidad de la información, sus usuarios potenciales, los estándares que servirán de comparación; entre otros. Dichos estándares pueden establecerse con base en valores históricos previos, metas planeadas o resultados de unidades de análisis similares a las evaluadas (a nivel nacional o internacional).

Los indicadores que miden costos, volumen de productos e incluso impacto son sencillos de diseñar en comparación con los indicadores de calidad, cuyo concepto es difícil de operativizar y más aún de establecer sus estándares. En este sentido se han utilizado cuatro formas de construirlos: 1) selección y seguimiento cuantitativo de algunos parámetros significativos, 2) aplicación de encuestas de satisfacción, 3) normalización de aspectos del nivel de calidad deseado (demoras, procedimientos, etcétera), y 4) compromisos o metas a alcanzar en términos de calidad (internos a la administración o difundidos públicamente) (Abate, 2000).

Además, los indicadores directos no siempre son posibles, por lo que a menudo hay que recurrir a sustitutos, con el riesgo de que la medición resulte afectada por factores extraños, sobre los cuales el programa puede no tener control alguno. También pueden presentar efectos no deseables como inducir sesgos que permitan

gestiones que cumplen con los indicadores descuidando los aspectos no medidos (calidad muy a menudo) o el logro global de los objetivos.

En el caso de los países en vías de desarrollo, la OCDE (1995) considera que el trabajo debe focalizarse hacia la selección de indicadores-clave en lugar de intentar una excesiva cobertura de los aspectos del programa, sin tener claridad sobre sus interrelaciones y posibilidades de uso.

Los organismos multilaterales suelen solicitar que los indicadores se apliquen siguiendo la orientación de su *marco lógico*: objetivo general, objetivos específicos, metas, indicadores, fuentes de información y resultados. La Comisión Económica para América Latina (CEPAL, Sanín, 1999) y el BID (1997) proponen usar una matriz de cuatro filas (objetivo superior, objetivo específico, productos o componentes y actividades) y cuatro columnas (resumen narrativo de objetivos, indicadores para medirlos, medios de verificación de los anteriores y supuestos de desarrollo). En el caso del BM (2004), se está promoviendo un esquema gubernamental mediante Administración Basada en Resultados (ABR) para la asignación de recursos presupuestales, que incorpora un sistema integrado de monitoreo y evaluación (M&E). Su marco lógico se compone de indicadores para: 1) monitoreo de desempeño continuo: de entrada, actividades y salida (resultados de corto plazo), y 2) evaluación de impacto (de largo plazo).

Los sistemas ABR han sido adoptados por diversos países de la OCDE (Nueva Zelanda, Reino Unido en la década de 1980, Canadá a principios de la de 1990, Estados Unidos a finales de esa década) y existen algunos esfuerzos por implementarlos en AL (Costa Rica, Uruguay, Colombia, Chile, Honduras, Bolivia). Insisten en que la información contenida en los indicadores debe ser entregada, no sólo a los distintos niveles de gobierno, sino a la sociedad civil en su conjunto.

Un problema relevante en América Latina para el funcionamiento de estos sistemas es el relacionado con la calidad de la información. Moreno (2007) considera que el poco éxito alcanzado por los procesos de evaluación se debe a debilidades estructurales de los sistemas existentes, insuficiente información sobre algunos sectores sociales, discontinuidad de los datos, limitaciones en su estandarización, poco desarrollo de bases de datos de rápido acceso, escasez de personal calificado en su manejo, falta de conexión confiable entre sus productores y usuarios, entre otros (Cardozo, 2009).

La información cuantitativa generada en los diseños anteriores, puede analizarse con *técnicas estadísticas*: análisis de datos (tipologías), análisis de series temporales, análisis multivariado (varianza y covarianza) y estimación econométrica sobre datos longitudinales agregados. Éstas también permiten dar tratamiento cuantitativo a la aplicación de encuestas de opinión, otro de los instrumentos de evaluación de resultados más empleado.

Finalmente, también existen técnicas cuantitativas diseñadas para hacer evaluación *a priori* que a menudo se adaptan para evaluar resultados, sustituyendo lo que antes eran previsiones por los datos históricos realmente obtenidos. Lógicamente, en los casos en que utilizan un criterio optimizador, éste se pierde porque generalmente las posibles alternativas de un proyecto no se ponen en práctica y por lo tanto, se carece de posibilidades aceptables de comparación. En su versión original *a priori* dichos métodos asumen las siguientes características (Cardozo, 1993):

El *análisis costo-eficacia* permite comparar programas y proyectos alternativos que persiguen un mismo objetivo, seleccionando el que presenta menor costo por unidad de logro (C/E). La eficacia suele ser difícil de cuantificar en forma directa, por lo que se tiende a sustituir su medición por indicadores indirectos, supuestamente correlacionados con los objetivos finales. La mayor limitación del método es no permitir la comparación de programas con objetivos diferentes o múltiples.

El *análisis costo-beneficio* pretende resolver el problema anterior homogeneizando los costos y los beneficios de toda índole (financieros, económicos, sociales, distributivos, etcétera) por medio de una unidad de medida común (generalmente las unidades monetarias), actualizando los flujos que se presentan en distintos momentos del tiempo, y comparando posteriormente ya sea las relaciones C/B de cada proyecto para seleccionar la menor, o bien, los valores actuales netos (VAN) para seleccionar el mayor de los positivos.

Este método ha sido criticado por: 1) las dificultades y arbitrariedades en que incurre para la cuantificación de efectos que no tienen un valor de mercado, lo que se realiza mediante la atribución de “precios sombra”; 2) la falta de criterios claros para determinar la tasa de actualización social y, en su caso, los factores de ponderación para los diferentes grupos de ingresos; 3) su tendencia a incluir sólo las variables de cuantificación más sencilla (económicas y financieras) a pesar de su objetivo manifiesto; y 4) su excesiva confianza en que un sofisticado análisis,

inaccesible para la mayoría de los interesados y basado en cuantificaciones discutibles, puede conducir a una selección óptima.

Los *análisis multicriterios*, especialmente el ELECTRA<sup>3</sup> diseñado por los franceses, han reaccionado frente a las críticas presentadas a la técnica anterior. También toman en cuenta objetivos múltiples y aplican valoraciones subjetivas y cualitativas en una escala preestablecida, mediante la participación de los grupos implicados en el problema. Proceden ordenando los programas en función de cada objetivo y negociando sus pesos relativos con los interesados o sus representantes. Son de carácter no optimizador, limitándose a buscar la solución más satisfactoria a partir de una simulación cuantitativa del regateo social que permite identificar el subconjunto de programas que presentan las mejores calificaciones y que al mismo tiempo carecen de fuertes oposiciones.

Finalmente, los *criterios* utilizados en las mediciones cuantitativas giran fundamentalmente alrededor de la eficacia (logro de objetivos), la eficiencia (logro de resultados al menor costo posible), la efectividad (impacto, incluyendo externalidades positivas y negativas) y la equidad (distribución de los objetivos logrados en relación directa con las necesidades existentes). Recientemente se han agregado otros como la sostenibilidad de los programas. Algunos autores comienzan a mezclarlos para hablar de efectividad sostenible, eficacia viable, eficacia eficiente, etcétera (Hintze, 2003). Cabe aclarar que cuando se habla de logro de objetivos, y principalmente de costos, no se hace referencia sólo a los de carácter económico, sino también a los sociales y ambientales que permiten una conceptualización adecuada de los criterios anteriores, comúnmente reducidos a la productividad o la rentabilidad financiera.

Es importante señalar que han sido muy pocas las evaluaciones que efectivamente han utilizado el criterio de equidad. Esto puede explicarse, en parte, por problemas metodológicos, y por la tendencia preponderante dentro de la teoría económica de no incluir en su análisis los problemas vinculados con la distribución del ingreso, cuyas consideraciones valorativas quedarían fuera del alcance de la ciencia positiva (Weisbrod, en Layard, 1978). Por otra parte,

<sup>3</sup> Electra es la traducción de la sigla francesa Electre, que corresponde a Elimination et Choix Traduisant la Réalité, que traducido al español significa Eliminación y elección que traduce la realidad.

como señala Rosenbaum (2002), el problema de la desigualdad nunca ha sido preocupación central de los gobiernos; sin embargo, en años recientes, a nivel mundial,<sup>4</sup> y tanto en Estados Unidos,<sup>5</sup> como en los países de América Latina<sup>6</sup> y México,<sup>7</sup> en particular, se incrementaron los niveles de desigualdad en forma considerable, por lo que este criterio debería ser considerado explícitamente en la evaluación, lo que resulta de particular importancia a la investigación en curso. Nos referimos tanto a la medición de la equidad a nivel regional, por grupos sociales, etáreos, de género, etcétera.

A manera de cierre de este apartado, puede decirse que la exposición previa muestra que la medición cuantitativa de resultados e impactos no constituye una forma óptima de identificar la efectividad de las políticas y tomar decisiones sobre su finalización o continuación, con o sin modificaciones. Para Nioche (1982), limitarse al paradigma experimental implica renunciar al aprendizaje de efectos no previstos, objetivos ocultos, teorías implícitas, etcétera, y querer imponerlo como criterio fundamental de cientificidad conduce a pseudoexperimentaciones sin sentido o al empobrecimiento del campo del análisis.

El Centre d'Études sur les Réseaux de Transport et l'Urbanisme (CERTU, 2002) comparte la idea de que las evaluaciones exclusivamente cuantitativas resultan reduccionistas, debido a la eliminación de los aspectos no cuantificables y sus interrelaciones, pero reconocen su fecunda contribución para fundamentar juicios de valor, elaborar estudios sintéticos y una batería de indicadores bien definidos, con plena conciencia de sus límites.

Por otra parte, en el análisis de las políticas públicas, los métodos cuantitativos no son siempre usados con la neutralidad que aquí han sido expuestos, sino

<sup>4</sup> Según la Organización de las Naciones Unidas (2005), el coeficiente de Gini mundial fue de 0.63, donde la renta del 20% de las personas más ricas supera en 28.7 veces la del 20% más pobre.

<sup>5</sup> Según datos retomados del BM, el coeficiente de Gini de Estados Unidos en el año 2000 era de 0.408, y el 19% más rico de la población superaba en 15.9 veces la renta del 10% más pobre.

<sup>6</sup> América Latina es la región más desigual del globo. Con datos de CEPAL 2004, su coeficiente de Gini era de 0.52, mientras los de África, Asia y los países de la OCDE alcanzaban 0.48, 0.40 y 0.33 respectivamente.

<sup>7</sup> México presenta uno de los índices de Gini más alto de América Latina (0.517), cercano al de Paraguay (0.52) para 2008, según BM y ONU.

que a menudo son manipulados para justificar y legitimar decisiones adoptadas con base en otros intereses (pseudoevaluaciones), en cuyo caso el modelo incrementalista de Lindblom (1959) explica mejor estos procesos.

### **Métodos cualitativos**

Es necesario mencionar que algunos métodos mezclan elementos cuantitativos y cualitativos. Para Lascombes y Setbon (1996) las soluciones aplicadas para alejarse de las críticas sufridas por el enfoque experimentalista, han tendido a esta combinación, como forma de equilibrar los resultados: informaciones cuantitativas recogidas por mediciones comparativas en muestras contrastadas, con datos cualitativos obtenidos por la observación etnográfica (es el caso del modelo de M.Q. Patton retomado por la General Accounting Office de Estados Unidos).

Un método menos cuantitativo de hacer evaluación que los planteados en el apartado anterior (puesto que también reciben tratamiento estadístico), lo constituye la aplicación de *encuestas* que recogen información, percepciones y opiniones subjetivas de una muestra representativa de destinatarios de las acciones.

Algo similar sucede en cuanto al grado limitado de cuantificación del *método DELPHI*, que se caracteriza por recoger información de expertos, desconocidos entre sí, que envían sus opiniones por correspondencia, mismas cuya media y desviación estándar se analiza y reenvía, para ver si los expertos desean hacer correcciones a sus propuestas, volviendo a repetirse el procedimiento hasta lograr el nivel de consenso requerido. Monnier (1995) considera que la utilización de este método sólo consigue crear consensos artificiales entre los actores, que no sobreviven a la hora de discutir los resultados de la evaluación.

Otro ejemplo podría constituirlo el *benchmarking*, retomado de la administración privada por la cultura de la gestión y la medición del rendimiento del sector público, que compara los procesos desempeñados en una organización con el líder en materia de “buenas prácticas” de su ramo. Hace uso de indicadores cualitativos en los casos en que no es posible cuantificar las variables a comparar. Pitarelli y Monnier (2000) consideran que se trata de una nueva herramienta, aplicable a la evaluación operativa y la de impacto, que permite juzgar el éxito de un programa por comparación con otro que sirve de modelo, pero advirtiendo

que también puede conducir a decepciones, dependiendo de la importancia que se le atribuya en el proceso de evaluación.

Un último enfoque que resulta fructífero es el diseño de *escenarios prospectivos* en los que se combinan análisis cuantitativos de tendencias con la consideración de los posibles eventos cualitativos que pudieran introducir cambios en las mismas. Esto permite mostrar a los tomadores de decisiones con más claridad lo que probablemente sucederá en el futuro si la tendencia se mantiene sin modificaciones y cómo podría afectarse si se produjeran eventos que escapan a nuestro control o si realizamos algún tipo de intervención intencional.

Los métodos estrictamente cualitativos han surgido, en su mayoría, de la crítica a un enfoque cuantitativo, de extracción generalmente positivista, que supone la existencia de rasgos universales y regularidades. Así, las ciencias sociales han pasado del estudio de grandes procesos estructurales (macro), con metodologías cuantitativas, a la construcción social de significados en las comunidades locales, con enfoque cualitativo. Los últimos enfatizan el análisis de las diferencias, el estudio de lo individual (micro), de sus motivos, significados y emociones, de las dimensiones sociales, culturales y políticas en que se desarrollan, así como la vida de los grupos en sociedad, sus acciones e interrelaciones.

Sin embargo, las críticas al cuantitativismo y las nuevas propuestas no han bastado para que se constituya en un enfoque alternativo, por lo que aún es necesario seguir trabajando en la fundamentación de su confiabilidad y validez, y en la búsqueda de nuevos paradigmas. Además, dichas críticas exigen un conocimiento profundo de los métodos cuantitativos, y lo más frecuente es la adhesión de ciertos investigadores al enfoque cualitativo por críticas mal fundadas hacia lo cuantitativo, o de plano, un rechazo ideológico o traumático a los “números”.

Un análisis serio y crítico de la potencialidad de los métodos cualitativos y de su utilización en la investigación empírica, sin descartar a los de índole cuantitativa, los define como “la conjunción de ciertas técnicas de recolección, modelos analíticos, normalmente inductivos y teorías que privilegian el significado que los actores otorgan a la experiencia” (Tarrés, 2001:16).

Dichas técnicas de recolección de datos se caracterizan por:

- La relación de involucramiento y empatía del investigador con el instrumento utilizado y con el objeto a estudiar (reflexionando, interviniendo y controlando

activamente el proceso), tratando de “observar, escuchar y comprender” al otro.

- Su orientación a la búsqueda de significados, ideas y sentimientos subyacentes o latentes en las descripciones obtenidas de las palabras o de las conductas observadas,
- Su prioridad a la narración y valoración de esa subjetividad, tratando de contextualizar las experiencias estudiadas y de interpretarlas teóricamente. No se preocupan por la representatividad de los casos analizados frente al universo total, por lo que no generalizan sus resultados, salvo con el carácter de hipótesis heurísticas.
- Su entusiasta preocupación por adaptar las concepciones y procedimientos sobre validez y confiabilidad para que el conocimiento generado cualitativamente sea aceptado y útil.

Entre los métodos cualitativos, los más recurridos son: la elaboración de estudios descriptivos de caso, biografías, historias de vida, la observación tradicional y participante, la investigación-acción, las dinámicas de grupo, los grupos de discusión, los paneles de opinión, los debates dirigidos, el análisis de textos, las entrevistas, etcétera.

Aplicados a la evaluación estos métodos permiten introducir el análisis de la motivación, intereses, significados, sentimientos, valores, actitudes, comunicación, conductas e interferencias culturales de los implementadores y beneficiarios de un programa. Intentan arribar a conclusiones sin medir los efectos de las intervenciones o bien complementan una medición previa con un análisis explicativo que reconstruye el proceso de su implementación para comprender los resultados alcanzados. Los tipos más utilizados en evaluación son la entrevista, el análisis de textos y los estudios de caso.

La *entrevista* o “encuentro hablado entre dos individuos que comporta interacciones tanto verbales como no verbales. No es un encuentro entre dos personas iguales puesto que está basado en una diferenciación de roles entre los dos participantes” (Pope, 1979, citado en Rubio y Varas, 1997). Se la ha definido como una situación construida con el fin específico de que un individuo pueda expresar, al menos en una conversación, sus referencias pasadas y/o presentes, así como sus anticipaciones o intenciones futuras (Kanh y Cannell, citado en Tarrés, 2001). A diferencia de la encuesta estadística, la entrevista utiliza un muestreo

de tipo teórico o intencional y se repite hasta alcanzar un “punto de saturación”, en la que el investigador considera que su continuación no agregaría información relevante. Cuando se dirige a informantes-clave intenta recoger información de aquellos que la poseen en forma privilegiada debido al papel que desempeñan en el proceso en estudio (directivos de programas, líderes de grupos, etcétera).

La entrevista también pueden asumir diversos tipos, de los cuales los más empleados en evaluación son: 1) semiestructurada, combinación útil para entrevistar personas que disponen de poco tiempo (directivos, burócratas, líderes, etcétera), 2) enfocada, en la que el investigador conoce de antemano la participación que ha tenido el entrevistado en algún evento de su interés y asume una posición directiva de la entrevista, conduciéndola hacia el asunto que constituye su foco de atención, pero dejando en libertad al entrevistado en la exposición del mismo, y 3) grupal, dirigida a un grupo focal sobre un tema compartido por todos, donde el investigador actúa como moderador.

El *análisis de textos* o *del discurso* es un método hermenéutico de análisis estructural, comprensión e interpretación de textos que estudia el contenido de los mismos (hermenéutica interna) y el entorno en que se producen (hermenéutica externa), de manera de “leer” incluso lo que no está escrito expresamente en programas, informes, etcétera. Busca reducir una cantidad importante de datos a una unidad de medida, mediante el empleo de la intuición del investigador. Entre sus tácticas figuran: contar, relevar patrones o temas recurrentes, buscar la verosimilitud o sentido de las palabras, agrupar, construir y utilizar metáforas, separar variables, relevar relaciones entre variables, entre otras (Miles y Huberman, citado en Díaz y Rosales, 2003).

Los *estudios de caso* pueden apoyarse tanto en la tradición cuantitativa como en la cualitativa y complementarse con métodos comparativos y experimentales. El caso se refiere al análisis de algo específico: entidad empírica o construcción teórica que es objeto de la investigación. Analiza el objeto individual a ser estudiado, manteniendo un enfoque holístico, que permite preservar la unidad del todo. Puede ser una forma de comprender mejor el caso particular o de ejemplificar un tema, pero también puede permitir mejorar la comprensión de un fenómeno general. En la elaboración de estudios de caso no se parte de un marco teórico o hipótesis, sino que se privilegia la descripción sobre la estandarización generalizadora, por lo que sólo es posible pensar en algún tipo de generalización de estos casos a partir de la acumulación de los mismos y la consideración de casos contrastantes, hasta

lograr la depuración del modelo. Cuando este estudio recae en una persona, puede tratarse de: una autobiografía, un relato de vida, una historia de vida, o un relato testimonial.

Los *criterios* empleados en estos estudios cualitativos, al aplicarse a la evaluación de programas, enfatizan aspectos vinculados a la calidad, satisfacción del usuario, congruencia, oportunidad y pertinencia, entre otros.

En relación con el análisis de la información cualitativa, el investigador trabaja con mucha libertad para dar sentido a los datos en estudio, sin expectativas o teorías preexistentes. Recientemente se han desarrollado programas de computación (ATLAS, NUD\*IST) para el manejo de información proveniente de entrevistas o de apoyo al análisis de textos, que permiten realizar conteos y clasificaciones complejas en poco tiempo.

Lógicamente los planteamientos cualitativos han desembocado también en una redefinición de los criterios de confiabilidad y validez, heredados del positivismo. Mientras los métodos cuantitativos gozan de amplia aceptación en estos términos, entre los cualitativos la preocupación ha aumentado, especialmente en años recientes, pero con un desarrollo lento, insuficiente y diversificado. No resulta fácil lograr un mínimo de rigor científico y concordancia aceptable de la estructura narrativa proporcionada por los informantes a partir de los hechos, tal y como se les presentan de manera cotidiana en su realidad. (Tarrés, 2001). Sherrard y Barrera (1995, en Tarrés, 2001: 89) plantean que la validez y confiabilidad de las entrevistas dependen de tres criterios:

- Legitimidad del investigador en el medio académico y ante su informante.
- Empleo apropiado de la técnica basado en su experiencia y entrenamiento.
- Calidad del *rapport* (aceptación o entendimiento mutuo) establecido con el entrevistado.

El problema que nos deja este tipo de planteamientos es que resulta muy difícil valorar estos criterios para quienes no participaron en el proceso, y por lo tanto, limitan la confianza para utilizar sus resultados como información secundaria. Sin embargo, se debe reconocer que, tampoco en materia de métodos cuantitativos podemos estar seguros en temas como la composición aleatoria de la muestra para garantizar su representatividad, o la forma en que se aplicó un cuestionario; por lo que parece que el único criterio que podemos tomar

en cuenta, en ambos casos, se refiere a la legitimidad y prestigio del equipo de investigadores y la institución que realizaron el trabajo.

En la evaluación de programas, los métodos cualitativos han sido especialmente utilizados por el enfoque pluralista. Monnier (1995) propone el modelo del *modus vivendi* o de negociación de acción. Es el que considera más adaptable a entornos cambiantes, ya que “la acción será juzgada positiva o negativamente, por una u otra parte, en virtud de que se hayan satisfecho sus intereses específicos en el marco de la acción realizada después de la negociación” (Monnier, 1995:106).

La legitimidad de este modelo se basa en el compromiso de las partes (gobierno, legisladores, administración pública, asociaciones profesionales, agentes económicos, expertos, investigadores universitarios y público en general), lo que, según el autor, reportaría mayores éxitos en programas públicos innovadores.

Los diversos actores mencionados pueden ubicarse en uno o más de los siguientes contextos: el de legitimación, compuesto por los responsables (jurídicos, científico-técnicos, operativos, etcétera) de la decisión; el de la acción programada, integrado por quienes han recibido un mandato de los miembros del grupo anterior para la ejecución del programa, que disfrutaban de cierta autonomía relativa; y el de reacción, conformado por todos los actores sociales que emiten un juicio sobre el valor del programa (evaluación). Esto explica la necesidad de muchas mediaciones para alcanzar un compromiso, siquiera inestable, entre partes tan heterogéneas. Por lo que el autor concluye que “cualquier evaluación da lugar a conclusiones relativas; estas conclusiones dependen del punto de vista del actor considerado” (Monnier, 1995:127-128).

Para Lascoumes y Setbon (1996), mientras que en los estudios experimentales es “el evaluador” quien carga el peso de entregar resultados rigurosos y creíbles (dimensión cognitiva); en la evaluación pluralista, ésta depende de una instancia con amplia participación de actores que representan visiones, valores e intereses diversos, creada específicamente para lograr, además de la dimensión anterior, la producción colectiva que garantice su aceptación como referencia (dimensión política). Se cuidan los aspectos metodológicos y se persigue una doble legitimidad: científica y política.

De esta forma, la evaluación pluralista renuncia a alcanzar conclusiones científicamente verificadas para negociar un compromiso aceptable con los

actores. Se trata de un intento colectivo de información y un proceso interactivo e iterativo para la elección de los criterios de la evaluación.

En el contexto francés este enfoque ha generado cierta polémica: Nioche (1992) considera que la propuesta anterior no constituye una evaluación, ya que el objetivo de ésta es informar al mundo exterior (no a los protagonistas, como señala Monnier), por lo que no puede basarse en criterios negociados, no sirve para responder si la política ha sido eficaz o no, si debe continuar o cambiarse. Nioche es partidario de una evaluación externa, que pueda complementarse con la interna, tratando de encontrar su racionalidad *a posteriori*, por una vía no optimizadora.

### **Comentario final**

Es importante resaltar que en la evaluación no hay métodos, cuantitativos ni cualitativos, que resulten siempre mejores que los demás; el o los métodos adecuados dependerán siempre de la situación a evaluar. Monnier (1995) plantea que éstos dependen de la perspectiva de quien la encarga (jurídica, técnico-económica o política), las características del programa (innovador o rutinario), del tipo de evaluación a realizar (de objetivos, de impactos, de proceso o pluralista), de la intención (endoformativa o recapitulativa), etcétera.

A pesar de lo anterior, se perciben énfasis diferentes y muy marcados de las dos tendencias identificadas al finalizar el capítulo anterior: los enfoques positivistas utilizan prioritariamente la experimentación y los métodos cuantitativos, mientras los construccionistas prefieren los estudios no experimentales de tipo cualitativo.

En una revisión hemerográfica realizada hace unos años (Cardozo, 1999), se comparó el papel desempeñado por los métodos cuantitativos en estas evaluaciones, en las dos fases del proceso de políticas públicas en que éstas resultan más útiles, con los estudios de carácter cualitativo que se publican sobre el tema, concluyendo que esta última perspectiva está siendo más productiva y probablemente, por diferentes razones, más aceptada por los lectores. La tendencia muestra más interés en el desarrollo de los segundos, que se supone predominarán en los próximos años, salvo que se logre mayor realismo en el desarrollo de modelos matemáticos aplicados.

En este trabajo se considera que los métodos cuantitativos constituyen una aportación valiosa al estudio de las políticas y programas públicos, siempre y cuando se tenga presente el conjunto de supuestos en que se fundamentan, y por lo tanto, las limitaciones que presentan, y que se les complemente con el estudio de los aspectos cualitativos, importantes en este campo, como son los problemas de negociación, consenso, legitimidad, satisfacción, calidad, etcétera. De esta manera se obtiene la adecuada integración de las racionalidades técnica y política, necesarias en el campo de las políticas públicas.

En particular, los métodos cuantitativos proveen los instrumentos que hacen posible contar con un panorama global y representativo de los logros y limitaciones de los programas que se evalúan; mismos que resultan necesarios pero no suficientes para la realización de una evaluación integral que pretenda explicar los resultados alcanzados.

Por su parte, los métodos cualitativos se consideran un complemento necesario de los cuantitativos para avanzar en el estudio de situaciones concretas y comprender aspectos subjetivos en mayor profundidad.

Una combinación de ambos tipos de métodos es considerada por House (1993, citado en Dopico, s/f) como la tendencia dominante; también es recomendada por el BM (2004a) para la realización de evaluaciones.



## CAPÍTULO IV

### Experiencias de evaluación en organismos internacionales y países desarrollados

Si bien se mencionan antecedentes de evaluación en la antigüedad (China en el siglo XX a.C. o Grecia en el siglo V a.C.), en la época moderna ésta reaparece en el siglo XIX en Inglaterra para la evaluación de servicios públicos y en Estados Unidos a inicios del XX en la evaluación del aprendizaje educativo.

A partir de 1980, la mayoría de los países ha intentado aplicarla. En la década de 1990, con la segunda generación de reformas del Estado, esto empezó a incrementarse debido a la necesidad de enfrentar la pérdida de legitimidad del aparato público, a los logros alcanzados por los países de la OCDE en la materia, al auge del NPM, y a la presión de actores sociales, entre otros.

En consecuencia, en el presente capítulo se procede a analizar el papel que han desempeñado en el tema los organismos multilaterales que actúan en el ámbito internacional y las experiencias de algunos de los países desarrollados con mayor avance en la materia, particularmente el caso francés debido a las similitudes que en su devenir guarda con el caso mexicano (centralismo, resistencia a la evaluación, falta de participación del Poder Legislativo, problemas administrativos, políticos, culturales y metodológicos, entre otros). Se agrega una visión global de los avances en los países europeos.

#### **Ámbito internacional: ¿control o evaluación participativa?**

Al inicio de la década de 1970, la evaluación pasó directamente de la burocracia federal de Estados Unidos a algunos organismos de la Organización de las Naciones Unidas (ONU) y a los bancos de fomento multilaterales con sede en Washington (Brown, 2001).

A mediados de la década de 1980, la evaluación logra su madurez y deja de tener un enfoque relativamente simple de monitoreo y control, para transformarse en un análisis más sistemático de procesos, orientado al aprendizaje, modelo asumido por la mayoría de las agencias de la ONU a inicio de la década de 1990.

Por esas fechas destaca el fomento a la creación de capacidad de evaluación por parte del Departamento de Evaluación de Operaciones del BM, quien intentó aplicar su Sistema de evaluación de resultados de la gestión, brindando asistencia a un banco regional de desarrollo y a seis países, entre los que figuran Brasil, Colombia y Chile.

Las agencias de cooperación se habían preocupado por identificar claramente ante la opinión pública los resultados de su asistencia y satisfacer sus propias necesidades de responsabilidad fiscal, excluyendo a los beneficiarios del proceso de evaluación, quienes la habían sufrido como un control impuesto, no como oportunidad de aprendizaje.

Esta situación insatisfactoria fue percibida por parte de los organismos, algunos de los cuales (BID, por ejemplo) modificaron su actitud, abandonando el criterio de “inculpación” que los caracterizó, y sustituyéndolo por el aspecto didáctico de la rendición de cuentas, basada en una cultura de la evaluación que reduciría las posibilidades de corrupción y reforzaría la responsabilidad de los funcionarios públicos.

Hacia fines de la década de 1990, Spitz (2002) identifica un nuevo énfasis en la evaluación, la medición de logros de corto plazo para la administración basada en resultados, que responde a los imperativos del *value for money* (originado en Reino Unido en el periodo de Margaret Thatcher), con el correspondiente debilitamiento de la orientación hacia el aprendizaje. A pesar de que, en teoría, estos enfoques podrían compatibilizarse, la tendencia entre los administradores de nivel medio ha sido regresar a los mecanismos de monitoreo y control, en el marco de un enfoque de “caja negra” de meras relaciones entre insumos y productos. Spitz considera que las unidades de evaluación (salvo unas 20 dedicadas a desarrollo rural) de las agencias de la ONU, que desde 1984 han formado parte de la UN Inter-Agency Working Group on Evaluation (IAWG), enfrentan crecientes conflictos de prioridad entre el enfoque cognitivo/de aprendizaje/de construcción de conocimiento y el de medición de logros de corto plazo, en materia de evaluación.

En septiembre de 1999, el BM y el FMI anunciaron un nuevo instrumento de política para enfrentar la pobreza en países de bajo nivel de ingresos, como precondition para lograr la cancelación de deudas de los mismos con los organismos mencionados: the Poverty Reduction Strategy Paper (PRSP). Manifestaron mucho énfasis en monitorear el impacto de los PRSP's, recuperando el enfoque participativo, tanto de la población como de los distintos ámbitos de gobierno.

Sin embargo, Foresti y otros (2002), miembros de prestigiadas organizaciones no gubernamentales (Oxfam, Save the Childrem), consideran que no pasó de ser una consulta superficial, ya que un genuino enfoque participativo implicaría abrir espacios para que puedan expresarse las voces que tradicionalmente no se han escuchado. En cambio, los sistemas de monitoreo de los países involucrados en los PRSP suelen estar a cargo de los servicios de estadísticas nacionales, que se concretan a hacer el seguimiento de un núcleo estándar de indicadores de pobreza (inscripción escolar, mortalidad infantil, nivel de ingreso, etcétera), a veces inadecuados para algunos contextos. Proponen su complementación con datos cualitativos que recojan las percepciones de la población en condiciones de pobreza y su evolución, sobre aspectos claves de la misma.

Al iniciar el nuevo siglo, a pesar de los esfuerzos realizados por medir impactos macroeconómicos y sociales (crecimiento, inversión en capital humano, redes de apoyo, etcétera) de los programas dirigidos a reducir la pobreza, y de los miles de millones de dólares destinados anualmente a ayuda para el desarrollo, Baker (2000) planteaba que, en organismos como el BM en el que laboraba, aún se sabía muy poco de cada programa o proyecto específico en términos de eficiencia en el uso de los recursos, impactos en la población y ventajas y desventajas de diseños alternativos.

Las razones señaladas para la resistencia a evaluar, por parte de los gobiernos, instituciones y administradores de proyectos, incluyen altos costos, prolongadas duraciones, complejidad técnica, resultados políticamente delicados, resultados no disponibles en el momento necesario, que no responden las preguntas adecuadas, basados en datos limitados en cantidad y calidad o realizadas con insuficiente rigor científico.

Sin embargo, la evaluación es especialmente necesaria en los países en desarrollo, donde los recursos son más escasos, y los problemas antes planteados pueden revertirse si se cuenta con el apoyo de las autoridades responsables del

programa, una planeación adecuada y oportuna y un presupuesto relativamente pequeño en comparación al costo total del programa o proyecto. Adicionalmente, resalta la importancia de contar con la participación de las partes interesadas desde el diseño de los objetivos y el enfoque de la evaluación, una buena comunicación entre los miembros del equipo durante el proceso y una combinación apropiada de métodos cuantitativos y cualitativos para lograr resultados de alta calidad, creíbles y generalizables. Implícitamente, Baker parece también reconocer problemas en la formación de los evaluadores profesionales, ya que el trabajo que publica es un manual dirigido a ellos, para hacer más accesible la información sobre la evaluación de impacto, en especial la dirigida al desarrollo social.

El BM manifiesta haber “reforzado sus esfuerzos para medir el impacto que tienen las intervenciones sobre los pobres” y estar dedicando varias de sus unidades a su realización, como el Grupo Económico de Investigación para el Desarrollo, departamentos regionales, etcétera ([www.worldbank.org/poverty](http://www.worldbank.org/poverty)). En su trabajo dedicado a la pobreza en México expresa: “Un reto decisivo para una administración pública eficaz es el establecimiento de un sistema de monitoreo y de evaluación (M&E) confiable y eficaz” (BM, 2004a:237).

Esta medición de impacto, que privilegia el largo plazo, también tiene sus opositores (Booth y Lucas, 2002, citado por Foresti), quienes sienten que se descuida la evaluación de los procesos y resultados iniciales e intermedios y la explicitación de los supuestos de los distintos actores, por lo que abogan por un enfoque que integre ambos aspectos, incluyendo la participación de los beneficiarios.

La enorme importancia que los organismos internacionales atribuyen a los procesos de evaluación puede percibirse realizando una rápida incursión en las páginas electrónicas de algunos de ellos. También es de destacar la promoción de la evaluación en el marco de la Unión Europea (UE): en 1988 se estableció la obligación de evaluar el empleo de los fondos comunitarios y en 1993 se expidió el Reglamento que establece que cada acción comunitaria será objeto de una evaluación *a priori*, concomitante y *a posteriori*. Actualmente la UE solicita evaluaciones externas sobre la cooperación al desarrollo comunitario a partir de la Unidad de Evaluaciones de EuropeAid, oficina ejecutora de la programación realizada por la Dirección General de Desarrollo de la Comisión Europea.

Ante el limitado papel real desempeñado por la población no organizada y los beneficiarios en los procesos de evaluación inducidos por los organismos

anteriores, surge la necesidad de hacer referencia a la participación de organizaciones de la sociedad civil, comúnmente conocidas como organizaciones no gubernamentales (ONG), y directamente vinculadas a los flujos financieros internacionales de fundaciones y otros organismos donantes.

En la última década, estas organizaciones han reforzado su papel como actores en materia de políticas de desarrollo social y comenzado a valorar la realización de evaluaciones para contar con un mayor conocimiento de las consecuencias de sus intervenciones. También han contribuido a defender el derecho a la información y a operacionalizar la rendición de cuentas en la práctica, al evaluar programas gubernamentales y estrategias de organismos internacionales, como el PRSP, para los que han demostrado contar con algunas ventajas importantes como la disposición de información independiente de la proporcionada por los gobiernos y otros actores de las políticas o su cercanía y sensibilidad para recoger el punto de vista de quienes viven en condiciones de pobreza.

Su participación en estos procesos, aunque numéricamente limitada a la fecha, les ofrece la oportunidad de lograr una mejor comprensión del fenómeno en estudio y de incidir en la toma de decisiones y el cambio en las políticas. Algunos de los factores que afectan la calidad de su participación serían: el grado de desarrollo democrático alcanzado; la relación entre sociedad civil y Estado; las diferentes y a menudo incompatibles expectativas de participación de gobiernos y sociedad civil; y el nivel de representación de las comunidades locales más marginadas.

Obviamente, sus evaluaciones también tienen algunas desventajas, como la desconfianza provocada por consideraciones de falta de capacidad técnica y habilidades específicas, representatividad y legitimidad.

En el marco de la evaluación de los PRSP, en AL, se destaca el papel crítico desempeñado, en el caso de Bolivia, por el Centro de Estudios de Desarrollo Rural y Agrícola, que monitorea los niveles de pobreza y desarrolla investigación para estudiar su conexión con los PRSP, a efectos de brindar información alternativa a la mostrada a partir de los indicadores de éstos, y generar un debate en las organizaciones sindicales y sociales, menos involucradas que las ONG (Foresti *et al.*, 2002).

A los limitados avances percibidos en materia de participación de las ONG en la evaluación externa de programas, éstas agregan la creencia de que muchas evidencias son ignoradas y manipuladas, por lo que discuten si la sociedad debe o

no continuar dedicándole tiempo y recursos a esta actividad. Otra es la situación en cuanto a la valorización del aprendizaje a partir de evaluaciones realizadas a sus propios programas y proyectos.

### **Países desarrollados**

Monnier y Barbier (citados en la bibliografía de políticas públicas urbanas [www.urbanisme.equipement.gouv.fr]) realizaron, a principios de la década de 1980, un análisis del estado de la evaluación en los 15 Estados miembros de la Comunidad Europea. Encontraron que los mayores avances en materia de publicación de informes de evaluación y en su utilización en el marco de negociaciones políticas entre los poderes Ejecutivo y Legislativo y entre actores nacionales y regionales ocurrían en los países nórdicos: Irlanda, Países Bajos y Reino Unido. En el resto, la ciudadanía y sus representantes aún no sacaban provecho de los trabajos realizados.

Guerrero (1995), retomando el trabajo de Ray Rist, presenta un estudio del estado de la evaluación en nueve países desarrollados. Sobresalen por sus condiciones favorables y su nivel de institucionalización: Estados Unidos, Canadá y Alemania; los Países Bajos muestran condiciones también favorables, pero con un menor desarrollo; Reino Unido, Francia, Noruega y Dinamarca presentan condiciones poco favorables y un desarrollo modesto; finalmente, Suiza cuenta con una institucionalización muy modesta de la evaluación, pero ésta es demandada por la sociedad civil, que además participa en sus procesos. Los resultados de las evaluaciones están totalmente accesibles al público en Estados Unidos, Canadá, Francia, Países Bajos y Suiza.

Por su parte, Thoenig (2002) refiere algunos resultados del estudio que hiciera en 1999 en unos 20 países de la OCDE. Considera que hay buenas posibilidades de que se adopten prácticas evaluativas si se respeta su naturaleza pragmática, orientada a la acción, a la generación de conocimiento aplicable y oportuno. Además, plantea que las experiencias más fructíferas serían las relacionadas con evaluaciones concomitantes y participativas, que constituirían un acompañamiento de los cambios en el comportamiento organizacional. Por otra parte, a la ya tradicional presión por la eficiencia, se agregaría actualmente, una menor capacidad del sector público para tratar políticamente los problemas

públicos, lo que abriría nuevas oportunidades a la evaluación de la acción pública, que ya no aparece centrada en el Estado, como sucedía con las políticas públicas. El futuro de la evaluación dependerá de su capacidad para adecuarse a los nuevos retos sociopolíticos y coadyuvar a la gobernanza.

La OCDE (2001) y la UE han impulsado reformas presupuestarias en los países miembros en el sentido de abandonar los controles centralizados, manejar su gestión con flexibilidad y privilegiar la obtención de resultados, por los que serán evaluados. Sin embargo, eso no ha hecho que la evaluación sea una realidad generalizada, por lo que en países como Italia la situación es la siguiente:

[...] la obligación de evaluar los programas financiados por la Unión Europea se traduce en un desarrollo reciente de la evaluación en diferentes dominios de la acción pública. Pero el progreso en la calidad de las evaluaciones y su utilización para fines de aprendizaje organizacional se enfrentan a fuertes resistencias en el seno de la administración y a la falta de competencias metodológicas. Este retraso en materia de evaluación tiene un origen a la vez cultural y organizacional (Stame, 1998:66).

En este apartado se trata de recuperar algunas experiencias, más avanzadas que las latinoamericanas, ya sea por el énfasis en sus instrumentos metodológicos (Estados Unidos) o por las dificultades que han debido superar para el desarrollo de una cultura de evaluación y la implantación de estos procesos en sus países, a menudo muy similares a las que se viven en México hoy. En el contexto europeo, revisaremos el caso del Reino Unido en que su establecimiento fue más sencillo, y el francés, en que se han enfrentado dificultades importantes. Agregamos el caso canadiense, debido a su paradójal desarrollo y el español debido a su cercanía histórica y cultural con México.

Se debe resaltar que la institucionalización de la evaluación en estos países ha asumido formas muy diversas. Pollit (1998) nos plantea que mientras en Estados Unidos ha seguido enfáticamente el modelo del mercado (es un negocio), en Francia se procedió a su institucionalización jerárquica y en el Reino Unido funciona como un conjunto de redes sectoriales.

*Estados Unidos: los dilemas de la experimentación*

Stufflebeam y Shinkfield (1987) realizan una útil periodización del desarrollo de la evaluación en Estados Unidos:

- 1904 a 1930: se le conoce como el periodo pre-tylerino, en que se extiende el experimentalismo, iniciado con base en la psicología por Alfred Binet para la evaluación del sector de instrucción pública. En 1912, Starch y Elliot desarrollaron las primeras evaluaciones de los servicios públicos educativos, en materia de enseñanza de las matemáticas, contando con un grupo que había recibido un entrenamiento especial y otro que funcionó como control.
- 1930 a 1945: la época tylerina, en que los análisis de programas de educación y salud pública fueron los más desarrollados, resaltando el diseño realizado por Tyler de un método de evaluación educativa, centrado en el nivel de logro de los objetivos.
- 1946 a 1957: la época de la “inocencia” o la irresponsabilidad social, posterior a la Segunda Guerra Mundial, en que la aplicación de técnicas de investigación social creció fuertemente. A fines de la década de 1950 se había consolidado en Estados Unidos y comenzaba a aplicarse en los países más desarrollados de Europa. Así, en Estados Unidos, la evaluación se desarrolla en el periodo llamado “sociedad de las reformas” y en Europa durante el “Estado de bienestar”. Los gobiernos fueron los principales demandantes de los procesos de evaluación y la sociedad fue el objeto de estudio que debía ser modificado.
- 1958 a 1972: la época del realismo. Estados Unidos utilizó la evaluación para impulsar sus programas de gobierno y controlarlos. Se diseñaron nuevos *tests* estandarizados a nivel nacional (en contraste con el método de Tyler que enfatizaba la diferencia en la aplicación de objetivos de un lugar a otro) y se utilizó el criterio profesional para valorar las respuestas y verificar periódicamente los resultados. Sin embargo, se empezó a comprender que la evaluación no estaba dando los frutos esperados, lo que se consignó en un histórico artículo de Cronbach (1963), que critica la falta de relevancia y utilidad de las evaluaciones realizadas y recomienda modificar las conceptualizaciones y métodos empleados. En 1965, con el lanzamiento de la Guerra contra la pobreza, la evaluación pasó a ser exigida por el Congreso

en otros programas sociales, además de los educativos. Posteriormente, se creó el National Study Committee on Evaluation.

- 1973 a la fecha en que escribieron los autores seguidos (1987): la época del profesionalismo, en que el rol de la evaluación se redefinió como herramienta de rendición de cuentas, a partir del proceso presupuestal, para medir el desempeño gubernamental en la consecución de objetivos y metas (Boyle y Lemaire, 1998). En este periodo la evaluación fue ganando reconocimiento en los organismos gubernamentales, institucionalizándose en las universidades y formando profesionales especializados (Ziccardi y Saltalamachia, 1997).

A partir de 1968, Estados Unidos comenzó a innovar con mucho rigor en el diseño de experimentos sociales que, en opinión de Monnier, fueron desilusionando a políticos y universitarios, por lo que en la actualidad el balance es bastante mediocre. Monnier (1995) identifica otra periodización en su desarrollo:

- De 1968 a 1975: es el periodo de grandes ambiciones innovadoras en manos de universitarios. A inicios de la década de 1970, el gobierno federal encargaba unas 200 evaluaciones anuales con un costo promedio de 100 000 dólares. Durante el mismo se desarrolló el famoso experimento del “Negative income tax”, con 4 800 familias y un costo de 132.7 millones de dólares de 1983. Posteriormente se llevó a cabo el “Great housing experiment” con 30 mil hogares y un presupuesto de 205.7 millones de dólares de 1983. En el último se hizo un especial esfuerzo por controlar los sesgos que terminaban por aparecer en todos los experimentos y que relativizaban sus resultados. El costo promedio de las evaluaciones de este periodo fue de 70 millones de dólares y la duración de los experimentos de 8 a 10 años.
- De 1975 a 1983: es la etapa de los avances modestos. Disminuyen los objetivos perseguidos por la evaluación y los medios aplicados, el tiempo se limita a unos tres años y los costos aproximadamente a la mitad. El tamaño de las muestras disminuye poco, pero su selección deja de ser guiada por principios científicos. El trabajo ya no se confía a los universitarios, se contratan agencias públicas y se cuestionan los resultados de las experimentaciones. La competencia entre el gobierno federal y el Congreso por contar con mayor número de evaluaciones llevó a que en 1977 se realizaran 1 315 evaluaciones.

- De 1983 en adelante: se trata de la etapa del reformismo prudente, con mayor retroceso del modelo experimental debido a la poca utilidad de sus resultados. El tamaño de las muestras vuelve a aumentar, pero disminuyendo sensiblemente los costos promedio (8 millones de dólares). El auge de la evaluación acabó durante el gobierno de Reagan, quien recortó los programas sociales de origen federal.

Actualmente este tipo de estudios es muy poco utilizado en Estados Unidos debido a los problemas presupuestales generados durante las crisis económicas, a la influencia de los conservadores que no se motivan con grandes innovaciones y a la desconfianza de políticos, universitarios y opinión pública frente a los costosos experimentos.

Monnier agrega que otra de las causas del fracaso de este enfoque de la evaluación se encuentra en sus postulados erróneos: *a*) racionalidad única y objetivos permanentes; *b*) relaciones lineales y unívocas entre el programa a prueba y sus efectos; *c*) representación utópica de una administración pública imparcial en la aplicación de la ley; y *d*) existencia de una solución única y uniforme, lo que provoca que la experimentación se justifique solamente cuando se pretende la generalización del programa a todo el país, si se comprueba que provoca los resultados deseados.

La Office of Management and Budget (OMB), responsable de la preparación del presupuesto federal, es quien ha coordinado en los últimos años la evaluación de programas gubernamentales. Por su parte, el poder Legislativo cuenta con el apoyo de cuatro agencias independientes del gobierno, que le permiten actuar como contrapeso: General Accounting Office (GAO), Congressional Budget Office, Office of Technology Assessment y Congressional Research Service. La GAO es, desde 1967, la responsable, a nivel federal, de evaluar políticas y programas públicos. La mayor parte de las solicitudes de evaluación la realizan los comités del Congreso o sus miembros individualmente; aunque también pueden efectuarse por solicitud del órgano responsable del programa o por obligación legal de la Sunset Legislation, que establece las instituciones, políticas y programas que deben evaluarse periódicamente.

En 1998 se aprobó la Federal Sunset Act que exige la revisión periódica de la eficacia y eficiencia de las agencias federales. Dichas evaluaciones son realizadas por una comisión *ad hoc*, compuesta de ocho congresistas y cuatro particulares,

que permanecen seis y tres años, respectivamente, en sus cargos. Trabajan en contacto con la GAO, la OMB, el Comptroller General y los comités del Congreso. Su producto es una recomendación dirigida al último sobre la conveniencia de continuidad o finalización de las agencias evaluadas.

Las condiciones sociales para el desarrollo de la evaluación son óptimas, en la medida en que la oposición cuestiona al gobierno desde el Congreso y ambas partes requieren evaluaciones para fundamentar sus posiciones. Además, existe un número importante de expertos en evaluación y recursos para desarrollarlas.

### *Reino Unido: un efectivo trabajo de convencimiento y capacitación*

Si bien desde el siglo XIX existen antecedentes de evaluación de servicios públicos, Saint-Clair (1993),<sup>1</sup> describe los primeros pasos que el Reino Unido dio para implantar la evaluación de políticas. Comenzó por crear un grupo mínimo (2 personas), con recursos suficientes para contratar consultores externos de corta duración; luego, definió los objetivos de la evaluación que se llevaría a cabo (mejorar la gestión, la toma de decisiones, la asignación de recursos, etcétera). Cubiertos estos dos aspectos, comenzó la tarea que cubrió varias etapas:

- a) Se hizo un diagnóstico del estado de la evaluación ya que, al igual que en México, todos los ministerios afirmaban hacerla, pero no pasaban de medir la eficacia de los servicios. Encontraron que los sectores más avanzados en la materia eran los vinculados al mercado: comercio, industria, empleo, entre otros. En materia social iban a la vanguardia salud y educación, pero sólo con evaluaciones de aspectos parciales.
- b) Se propuso cambiar la concepción de la evaluación: de la auditoría y la inspección, a un elemento constitutivo de la gestión administrativa. A diferencia de otros gobiernos que ubicaron a la evaluación en los Ministerios de Finanzas o las efectuaron por medio de despachos de auditores externos, se resolvió la aplicación de evaluaciones internas a los Departamentos.

<sup>1</sup> Director Adjunto de la Joint Management Unit.

- c) Se diseñó un plan estratégico, enfocando el trabajo a los 300 funcionarios con rangos de subsecretarios y responsables de los grupos de evaluación de políticas para convencerlos de la posibilidad y la utilidad de la evaluación. Éste se desarrolló también en diversas fases:
- Durante tres años, la evaluación se introdujo en el programa de formación de altos funcionarios (jóvenes de 30 a 40 años), con una duración de seis semanas, a tiempo completo, y organizó en cada ministerio una serie de 50 seminarios de media jornada para todos los altos responsables de la función pública. La misma incluía una exposición de una hora y luego un trabajo en equipo para intentar evaluar una política de su propio ministerio. Finalmente, se hacía un balance para analizar lo que podría mejorarse en materia de organización de la evaluación, diseño de indicadores-clave, etcétera. Contrario a lo que se temía, las respuestas a este proceso fueron muy positivas.
  - Posteriormente, se invitó a los Departamentos a proponer uno o dos estudios de caso en áreas que no habían sido sistemáticamente evaluadas, con el objetivo de aplicar los métodos y técnicas aprendidos. Se prepararon documentos de aproximadamente 10 páginas, mostrando los beneficios de la evaluación y las posibilidades de su aplicación en otros Departamentos. En paralelo, el Ministerio de Finanzas comenzó a exigir la evaluación como requisito para la aprobación de programas con gasto público.
  - Se dejó a los ministerios la decisión de la importancia que le atribuirían a la evaluación: los más avanzados crearon servicios especializados; los menos, la dejaron en manos de los mismos jefes del servicio, con el apoyo de una célula de expertos en metodología.
  - En 1988 se publicó un manual, "Policy Evaluation. A guide for managers", que se envió a los altos funcionarios, para el desarrollo de evaluaciones de políticas públicas. Sus principios eran presentados en forma muy simple, acompañados de un glosario con los principales términos especializados, a fin de desmitificar la evaluación y evitar los temores que la acompañan.
- d) En 1991, se hizo una nueva revisión de avances: algunos sectores habían integrado la evaluación a sus procesos; otros presentaban avances mínimos.
- e) En 1992, el Ministerio de Finanzas reforzó su importancia, al vincularla más con el sistema de asignación de recursos. En paralelo, trabajó con la Comisión

Parlamentaria de Control de las Cuentas Públicas para mejorar la presentación de la información al Legislativo.

En resumen, el gobierno británico logró avances importantes en el campo en 1980-1990, lapso en el que Margaret Thatcher impuso la evaluación de la eficiencia por medio del concepto de *value for money*, que incluye auditorías de desempeño, exámenes especiales y evaluación de programas a nivel microeconómico. En este enfoque el máximo evaluador es el mercado, compuesto por oferentes y demandantes; de ahí la tendencia a simularlo cuando no existe. Sus dificultades no fueron de orden metodológico, sino vinculadas con la resistencia de los responsables de programas para dedicar tiempo a esta actividad, en detrimento de otras consideradas más urgentes.

La Audit Comision, organismo independiente que evalúa en forma externa la eficiencia de los servicios que brindan las comunidades locales, implantó desde 1994 la aplicación de indicadores no financieros, los que deben publicarse anualmente y han provocado un importante interés entre los ciudadanos. Permiten realizar el *benchmarking* entre entidades similares y analizar la evolución histórica de los mismos.

En 1997, el gobierno británico adoptó un programa nacional piloto, *Best Value*, que combina el enfoque racional-objetivo de la evaluación de resultados e impactos con otro de tipo argumentativo-subjetivo, que enfatiza el aprendizaje organizacional. Esto debido a que la evaluación racional produce bancos de datos pero no genera información útil o cognición, no provoca cambios de comportamiento en la acción (Thoenig, 2002).

El Tesoro, encargado de la política presupuestaria y financiera, es quien incita a las administraciones a evaluar sus programas y difunde normas metodológicas para su desarrollo.

En el parlamento, el Public Accounts Committee cuenta con el apoyo de la National Audit Office (NAO) y tiene atribuciones para pedir cuentas a los responsables de programas, pero dispone de pocos recursos para cumplir esa función. La NAO está facultada para examinar la actividad de los departamentos ministeriales y de los órganos públicos desde el punto de vista económico, de eficiencia y eficacia. Cada año presenta alrededor de 50 informes del tipo *value for money* al Parlamento (Goldsworthy, 1998).

No existe ninguna autoridad dominante que controle el conjunto de actividades de evaluación, si bien, la influencia de las instituciones centrales de gobierno es importante. Tampoco hay un grupo de grandes firmas que concentre los recursos de evaluación, como en Estados Unidos (Pollit, 1998). Cuenta con una reducida disponibilidad de especialistas en evaluación y poca aplicación de las ciencias sociales a este campo, donde el componente político queda casi excluido. Sin embargo, se ha desarrollado una evaluación crítica en los centros de investigación que luchan por conservar su independencia, que pone en evidencia el carácter discriminatorio de los dispositivos oficiales, la reproducción y profundización de las diferencias entre los jóvenes, los efectos perversos de sus criterios en la asignación de recursos, entre otros (Freyssinet, 1998).

En un seminario realizado en 1997 en Francia se describía a la evaluación en Gran Bretaña como más desarrollada que en el país anfitrión en todos los niveles de la Administración Pública, disponiendo de recursos financieros y humanos más importantes y haciendo un uso más sistemático de sus resultados en la gestión. La misma es realizada mediante la conformación de redes de expertos sectoriales, que incluyen universitarios, profesionales y miembros de la Administración Pública, y no se limita a los estudios de *value for money*, utilizando una pluralidad de métodos cuantitativos y cualitativos, incluyendo el empleo de protocolos experimentales, aunque en forma menos frecuente que en Estados Unidos.

### *Canadá: ¿evaluación real o mero ritual?*

La evaluación de programas ha sido siempre monopolio del Poder Ejecutivo. Dio inicio en 1978 a cargo de la Oficina del Contralor General del Canadá, aunque existían antecedentes desde la década de 1960. Éste definía los estándares metodológicos, formaba parte del equipo que seleccionaba a quienes harían las evaluaciones, dirigía la formación de evaluadores, elaboraba guías y demás documentos de apoyo. Su reinado en la materia duró hasta su desaparición en 1995.

En 1987, Canadá era el país más avanzado en materia de evaluación (Saint-Clair, 1993). Cada ministerio contaba con una unidad de evaluación, responsable ante el ministro del ramo, que aplicaba evaluaciones a todo el conjunto, a

intervalos regulares, de forma que todos fueran evaluados en ciclos de cinco años. Para ello se realizaban entre 300 y 400 evaluaciones anuales.

Sin embargo, Lacasse (1993) tenía la hipótesis de que muy pocas se enfocaban a analizar sus aspectos más trascendentes: la pertinencia de la política, la necesidad de su permanencia o las propuestas de cambios, etcétera. Un examen de los informes demostró que se dedicaban sólo unas pocas frases rituales al tema central y la propuesta de algunas “pistas a explorar”. En teoría, los evaluadores sabían que debían revisar la razón de ser de los programas, pero todos consideraban que en la práctica era imposible realizarlo con rigor, inútil intentar convencer a los decisores de su importancia, o incluso ilegítimo, debido a las restricciones que pesaban sobre la actuación del ministerio. También afirmaban que al evaluador no correspondía realizar las propuestas, sino a los consejeros en gestión, además de que nadie se las solicitaba.

Se encontró que las “verdaderas” evaluaciones solicitadas por el Contralor no se llevaban a cabo, siendo sustituidas por evaluaciones “realistas”, más débiles, demandadas por los propios ministerios. Además, el evaluador debía cuidar que su trabajo no excediera los límites del propio ministerio y evitar la formulación de propuestas de cambios estructurales que implicaran altos costos potenciales a su cliente.

Se llegó así a una ritualización de la evaluación que no podía cuestionar la existencia misma del programa a cargo del ministerio que solicitaba su evaluación, sólo podía proponer mejoras marginales. Se trataba, entonces, de una paradoja: el país contaba con una amplia institucionalización de la evaluación, pero la forma en que realmente se aplicaba, defendía los intereses ministeriales y del evaluador, quien optaba por colaborar en lugar de cumplir su papel de agente de cambio. Complementariamente, había muy pocos trabajos sobre la utilización de los resultados de la evaluación.

Lacasse concluyó que, a pesar de todo lo escrito en materia de metodología de la evaluación, ésta seguía siendo un instrumento manipulable, por lo que se requería de una instancia capaz de funcionar como contrapeso de los intereses sectoriales y organizacionales. Por otro lado, consideraba imposible intentar evaluar sistemáticamente todas las políticas y hacerlo en profundidad (dilema entre cantidad y calidad), tanto por falta de recursos como por imposibilidad de que el sistema político-burocrático digiriera sus resultados, por lo que propuso contar con algunas evaluaciones efectuadas a fondo, con fuertes dosis de análisis

político, aprovechamiento de oportunidades, etcétera, sólo en los casos en que se requiriera de cambios sustantivos y se estuviera dispuesto a realizarlos, mientras el resto continuaba con el mero ritual.

En julio de 1994 comenzaron a aplicarse tres nuevas políticas, entre las que sobresale el Examen de los Programas, tendiente a una gestión por resultados que pretende saber: 1) si la política o el programa está de acuerdo con las prioridades del ministerio y responde a las necesidades del momento; 2) si logra sus objetivos de acuerdo con su presupuesto y sin imponer repercusiones no deseadas; y 3) si utiliza los medios más apropiados y eficientes para alcanzar sus objetivos. En el marco de esta política principal se aplican también las políticas de Evaluación y de Verificación Interna. Dentro de la primera se prevé el nombramiento en cada ministerio de un gestor superior, responsable de realizar las evaluaciones, independiente de las actividades evaluadas.

A este examen de programas se atribuye la disminución del gasto en 12% entre 1993-1994 y 1998-1999, y su concentración en gasto social. Sin embargo, su éxito para modificar la tendencia previa es sólo parcial, persistiendo múltiples problemas, entre los que sobresale 60% de las evaluaciones sin medición de impactos, debido al énfasis exclusivo en la satisfacción del cliente, o, la ausencia de análisis de su pertinencia, por concentrarse sólo en su procesos de implementación (Turgeon, 1998).

Este último autor mencionaba que la evaluación adquiriría tanto la forma *prêt à porter* (positivista) en el caso de programas limitados, de un solo objetivo y experimentales, como su confección *hecha a la medida* (constructivista) en el caso de las políticas interministeriales; sin que ninguna intentara imponerse sobre la otra forma. Desde 1994 se ha tratado de involucrar a los grupos de interés.

El caso del programa de acción comunitaria para los niños tuvo la particularidad de conjuntar los dos enfoques anteriores para dar una respuesta más amplia a sus requerimientos: responder con seguridad a la generalización estadística de sus posibles impactos y adaptar las evaluaciones a las preocupaciones cualitativas de quienes gestionan los proyectos, sobre todo a nivel regional.

Turgeon planteaba que la tendencia era que la evaluación continuara en manos del Ejecutivo (ante un Legislativo absolutamente ausente) y que, si no se introducían cambios, se continuaría sin conocer realmente el impacto de los programas, como había sucedido en los últimos 20 años.

Actualmente, Canadá orienta su gestión por resultados, reemplazando controles *a priori* por la evaluación *a posteriori*, con base en la Ley sobre la Administración Pública del año 2000. Este enfoque se acompaña de una amplia participación de los funcionarios centrada en el aprendizaje de la utilización de los resultados en la toma de decisiones, de la elaboración de un balance de la utilidad de la evaluación realizada, de una rendición de cuentas obligatoria a la ciudadanía y de un reforzamiento del rol de la Asamblea General en la evaluación de la gestión.

Los aspectos considerados en la evaluación son: pertinencia de las políticas y programas al momento de su evaluación; coherencia entre actividades, resultados y misión; nivel de logro de los objetivos; impactos positivos y negativos; y, si corresponde, los medios cómo podrían alcanzarse los objetivos en forma más eficaz, mediante ajustes al programa actual o con la introducción de otro nuevo. Los criterios empleados en su evaluación son, entonces, la pertinencia, la eficacia, la eficiencia y el impacto. Las cualidades que deben integrarse en la evaluación incluyen: la validez, la fidelidad a la situación, la credibilidad y la utilidad.

Cada ministerio u organismo decide qué programas serán evaluados y mediante qué procesos, salvo que se haga por iniciativa del Consejo del Tesoro. También pueden realizarse a nivel interministerial y gubernamental en su conjunto.

Las organizaciones que fomentan la evaluación de programas y la rendición de cuentas son el Parlamento, el Ministerio de Finanzas y el Consejo del Tesoro. Existen muchos especialistas dedicados a la evaluación, provenientes de las ciencias sociales.

#### *Francia: en búsqueda de un modelo adaptado a su idiosincracia*

La evaluación aparece en Francia, en la década de 1980, con la expectativa de mejorar la gestión pública y aumentar la democratización de la toma de decisiones. El antecedente inmediato de la evaluación en este país se encuentra en la década anterior, en que el Ministerio de Finanzas promovió la evaluación *a priori*, mediante la RCB, similar al PPBS de Estados Unidos.

El periodo 1980-1985 se ubica como de reflexión sobre el tema. Nioche (1982) identificaba problemas sociopolíticos, administrativos y metodológicos

que dificultaban su avance. Este autor señalaba, en 1983, al hacer un balance de la situación, que Francia realizaba muchas evaluaciones por demanda institucional, pero estaba muy atrasado en cuanto a la demanda social y ciudadana, tal cual se presentaba en Suecia, Dinamarca o Canadá. En 1983, se llevó a cabo un importante Coloquio Internacional sobre la Evaluación de las Políticas Públicas, que sirvió para revisar los avances realizados y dejó planteadas muchas inquietudes.

En cambio, en el siguiente decenio, se incrementa notablemente el desarrollo de la evaluación. En 1986, se produce el *Informe Deleau*, realizado con Nioche, Pens y Poinard, a solicitud del gobierno socialista (Deleau y otros, 1986). En él se define a la evaluación de políticas públicas como “identificar y medir sus efectos propios”, con dos características principales: a) el énfasis en los efectos de la acción pública hacia el exterior, a diferencia del control tradicional enfocado en los procesos internos, y b) su voluntad de medición cuantitativa, sumaria, que evitara que su desarrollo tardara periodos muy extensos. Ésta se basa en dos principios deontológicos: la autonomía financiera e institucional del evaluador que garantiza su objetividad, y la replicabilidad posible de los resultados por otro evaluador que utilice la misma metodología. Se propone un proceso a cargo de alguna instancia habilitada al efecto (gobierno, parlamento, jueces, etcétera), sin participación ciudadana. Monnier (1995) define esta evaluación como de carácter “gerencial”, preocupada por la eficiencia y la optimización de los recursos presupuestales y olvidada de la dimensión política.

En 1988 se discuten dos alternativas: crear un consejo nacional de evaluación, de carácter independiente, o bien un consejo interministerial de evaluación, de tipo gubernamental.

En 1989, se presenta el *Informe Viveret* al gobierno de Francois Mitterrand, que enfatiza el control democrático. En él, la evaluación de una acción pública se transforma en “emitir un juicio sobre el valor de esta acción”, que sólo los tribunales y los representantes electos tienen legitimidad para realizar, con excepción de los casos de autoevaluaciones. Dicho juicio debía fundamentarse en cuatro garantías para que la evaluación cumpliera con su responsabilidad democrática de movilización de la capacidad crítica de la ciudadanía: independencia de criterio (salvo en las autoevaluaciones), competencia y rigor, transparencia en el enunciado de los métodos y criterios aplicados, así como en los resultados obtenidos, y pluralismo de opiniones en lugar de la persecución

de una verdad única. A esto se agrega, salvo casos excepcionales, la regla de publicidad de los resultados. Se trata de una evaluación destinada a enfrentar “la debilidad del Poder Legislativo”, “el monopolio de la pericia en manos de la Administración Central” y “el déficit de las formas de organización democráticas” (Viveret, 1989; citado por Monnier, 1995).

La intención de esta evaluación era también que el decisor actuara con conocimiento de las consecuencias de sus actos, o sea que, buscara combinar su aportación al reforzamiento de la democracia y a la eficacia administrativa. Monnier (1995) la define como una evaluación “democrática”, por contraste con la sostenida en el *Informe Deleau* y con la evaluación “pluralista”, impulsada por él mismo durante el debate gubernamental previo a la presentación del *Informe Viveret*. Como consecuencia de este informe, se optó por la creación del Comité Interministerial de Evaluación (CIME), acompañado de la integración del Fondo Nacional para el Desarrollo de la Evaluación (FNDE).

El CIME estaba compuesto de cinco miembros permanentes y de los ministros implicados en el tópico evaluado. Los primeros eran propuestos por el Consejo de Estado, la Corte de Cuentas, el Mediador Nacional, o el Consejo Económico y Social. Este último es un cuerpo consultivo que representa a las organizaciones nacionales socio-profesionales, que puede sugerir el límite de ciertos financiamientos (OCDE, 1995). El CIME decidía sobre el inicio de la evaluación, aprobaba su metodología, acordaba su financiamiento y recibía el informe. El FNDE cubría 50% de los costos de evaluación, mientras el resto lo pagaban los ministerios.

En menos de dos años se crearon unas siete instancias permanentes, lo que hace a Duran y Monnier (1992) considerar que Francia entró al tema de la evaluación con mucho formalismo y una cierta precipitación. En la práctica, sólo el gobierno aplicó evaluaciones, bajo la presión de los acuerdos comunitarios de Bruselas y en el contexto de las reformas de centro-izquierda en la Administración Pública.

En 1990, el gobierno de Michel Rocard emitió el decreto número 90-82 que creó formalmente el sistema nacional de evaluación de políticas públicas, basado en principios de transparencia, pluralismo, independencia y rigor, cuya definición fue: “Investigar si los medios jurídicos, administrativos y financieros puestos en práctica permiten producir los efectos esperados por esta política y alcanzar los objetivos que se han señalado” (traducción libre). Como puede verse,

se retrocedió a poner énfasis en la evaluación *a priori*, olvidándose del pluralismo y la participación del Parlamento. La evaluación quedó a cargo del CIME, el FNDE y de un Consejo Científico de Evaluación (CCE) que garantizaría el pluralismo metodológico, la independencia de las evaluaciones y desempeñaría un rol pedagógico importante. Sus miembros, once personalidades independientes, con distintas experiencias profesionales y supuestamente especialistas, serían designados por seis años por el presidente de la República y presididos por el primer ministro.

Una vez que el CIME, organismo con mayor poder que el CCE, decidía las políticas que habrían de evaluarse, se integraba una instancia específica, compuesta de 10 a 12 personas, entre funcionarios, especialistas, investigadores universitarios y responsables de programas, que diseñaban la evaluación (sus métodos, técnicas, fuentes de información, etcétera). El CCE debía emitir un comentario inicial sobre la calidad de la propuesta y finalmente sobre la evaluación realizada, misma que se publicaba junto con ella. También debía hacer un reporte anual sobre las actividades de evaluación desarrolladas y los avances en materia de formación de evaluadores. En realidad, ninguna de estas actividades se cumplió en los primeros años de su creación.

En este periodo proliferaron las instancias de evaluación o “comités de sabios”, según Monnier (1995), cuya independencia y neutralidad garantizarían la legitimidad de la acción del Estado. Además de las ya nombradas a nivel intersectorial nacional (CIME, CCE y FNDE), se crearon otras de carácter permanente: una instancia de política sectorial (Comité Nacional de Evaluación de la Política de la Ciudad), otra vinculada con la aplicación de una ley (Comisión Nacional de Evaluación del Salario Mínimo Interprofesional), y dos relacionadas con programas públicos (Comité Nacional de Evaluación de la Investigación y Comité de Evaluación y Prospectiva de Nuevas Calificaciones Profesionales). En la práctica, estas estructuras no se componían sólo de expertos, sino que integraban miembros del Consejo de Estado y el Tribunal de Cuentas, lo que las mantenía fuertemente vinculadas al tradicional control administrativo y con riesgo de ser absorbidas por el sistema que se proponían modificar, sobre todo porque el parlamento quedaba totalmente al margen y no había desarrollado su especialidad en materia de evaluación (Muller, 1994).

También fue publicada una *Guía de Evaluación*, que sistematizaba la experiencia de la Misión de Evaluación y Prospectiva, del Ministerio de

Equipamiento. Además, entre 1988 y 1993, se llevaron a cabo cinco seminarios para discutir temas como la relación de la evaluación con la administración, la investigación, la decisión y el juicio valorativo, así como, el rol del evaluador. Por su parte, el CCE definió los cinco criterios que guían la evaluación de políticas en Francia: pertinencia de los objetivos a los problemas que el programa intenta resolver, coherencia entre su concepción y su implementación, eficacia en el logro de sus objetivos, eficiencia en la relación costo-beneficio, efectividad o impacto de las consecuencias globales de la política. Para Crozier (1989), siguió siendo una evaluación tecnocrática, sin participación de los demás competentes, los usuarios y administrados, cuyo grado de satisfacción se ignora.

Alrededor de 1995 se produce la etapa de maduración de la evaluación, de balance de lo realizado y también de cierto desinterés por ella. En la práctica se habían presentado sólo 27 proyectos de evaluación,<sup>2</sup> de los cuales sólo 10 se finalizaron, tardando para ello entre 2.5 y 4 años. Ninguna de dichas evaluaciones fue examinada en su fase terminal y se reconoció oficialmente que no fueron utilizadas para retroalimentar la toma de decisiones gubernamental. Las razones de lo anterior fueron: técnicas, porque se encontraron problemas vinculados con la definición de los objetivos de la evaluación, su carácter meramente descriptivo, dispersión, heterogeneidad y pérdida de datos, entre otros; y políticas, ya que se enfrentaron problemas de credibilidad en la dinámica del trabajo colectivo y las negociaciones realizadas.

Los sectores que más avanzaron en la evaluación de sus políticas fueron: desarrollo regional y urbano, equipamiento, salud y educación. Monnier (1995) critica el tipo de evaluación gerencial y “apolítica” que se ha aplicado en la educación universitaria, la investigación y la salud, donde más que entre expertos en evaluación, la discusión se produce entre “antiguos pares” universitarios, investigadores o médicos, enfoque que también subyace en los proyectos aprobados por el CIME. En los primeros dos casos se formó un Consejo de Evaluación Universitaria y de Investigación que, en opinión de Thoenig (1996), obligó a la realización de discusiones argumentadas y a una evaluación con énfasis en los procesos, más que en los impactos.

Por otro lado, se percibía una tendencia a no querer perjudicar a nadie con los resultados de las evaluaciones, lo que se tradujo en información poco útil y una

<sup>2</sup> Fueron 10 evaluaciones en 1991, siete en 1992, seis en 1993, cuatro en 1994 y 0 en 1995.

permanente tentación por autoevaluarse. Además, el parlamento francés no se había mostrado motivado por capacitarse en materia de evaluación para asumir una misión de control del poder Ejecutivo. Sólo contaban con un Servicio de Evaluación de las Elecciones Científicas y Técnicas, con mínimas atribuciones y pocos recursos.

Nioche (1993) resumía los problemas enfrentados por la evaluación a partir del decreto de 1990 en los siguientes: concepción limitada de tipo gerencial, estructura administrativa compuesta por tres organismos controlados por el Ejecutivo y ausencia de una institución independiente especializada en evaluación, falta de participación del parlamento, procedimientos complejos en que los roles y las responsabilidades no estaban claramente delimitados, falta de desarrollos metodológicos propios para enfrentar el rechazo relativo de los extranjeros, trabajo disperso y de poco rigor, limitada utilización de sus resultados en el debate público y la toma de decisiones administrativa, entre otros. Consideraba que la institucionalización de la evaluación de la acción pública requería de una reforma del Estado, que modificara el funcionamiento del parlamento y los cuerpos de control. Troza (1993) agregaba la existencia de un incremento de procedimientos y un déficit de aprendizajes, así como resaltaba la falta de conexión entre la evaluación y la administración interna.

Por su parte, Lascoumes y Setbon (1996) describían las evaluaciones realizadas como: cuidadosas en la selección plural de actores, con fuerte presencia de la administración y los cuerpos de inspección, una cultura dominante de inspección y auditoría, un déficit generalizado de profesionalización, una amplia autonomía de las instancias de evaluación, y generalmente, la presencia de un actor administrativo, que demandó la evaluación, forma parte de dicha instancia, y trata de desempeñar un rol central. Setbon (1998) agrega una débil implicación de los actores políticos que impide que las evaluaciones se traduzcan en un reposicionamiento del tema en la agenda y en su transformación, mientras que se manifiesta un amplio interés de los administradores que operan programas, un avance en materia de esclarecimiento de problemas y de aprendizaje social bajo el concepto de evaluación formativa.

En 1996 se crearon dos Oficinas Parlamentarias. La primera, referida a la Evaluación de políticas públicas, para enfrentar la necesidad de evaluar, retrospectiva y prospectivamente, la acción gubernamental y también la del propio Parlamento; todo esto, nuevamente, en el marco de las directivas

adoptadas por la Comunidad Europea. La segunda, dedicada a la Evaluación de la legislación, de composición estrictamente política, con menos poder que la anterior, ya que si bien puede recurrir a expertos, carece de un ordenamiento administrativo y financiero que facilite su misión. Esta decisión permitió restablecer un cierto equilibrio en los contrapesos entre poderes del sistema francés, con cuatro riesgos importantes: una evaluación meramente política (no objetiva, que defienda intereses partidarios), sin suficiente garantía científica, que dispusiera de medios limitados, y apoyada por una voluntad política incierta. Ese año se publicó también una *Pequeña guía para la evaluación de políticas públicas*.

Por último, hacia fines de la década de 1990, Francia vivió una cierta consolidación y madurez institucional de la evaluación, acompañada de un relanzamiento de la misma a nivel gubernamental y parlamentario. En noviembre de 1998, se produjo un nuevo decreto del presidente de la República (número 98-1048), que creó el Consejo Nacional de Evaluación de Políticas Públicas (CNE) y manifestó que “la evaluación de una política pública, en el sentido del presente decreto, tiene por objeto apreciar, en un cuadro interministerial, la eficacia de esta política, comparando sus resultados a los objetivos asignados y los medios empleados” (traducción libre). Se trata del regreso al concepto de evaluación *ex-post* para el mejoramiento de la eficacia de las administraciones públicas, basado en la aplicación de procedimientos más rápidos de evaluación (máximo de 18 meses, salvo casos excepcionales). El decreto busca un equilibrio entre la participación plural de los interesados (*stakeholders*) al estilo americano y el control, bien francés, del proceso por la administración. Se insiste en la transparencia y publicidad de los resultados, así como en una nueva apertura del CNE a los representantes parlamentarios y miembros del Consejo Económico y Social.

El CNE, compuesto de 14 personas de tiempo parcial, nombradas por el gobierno por periodos de 3 a 6 años, está encargado de definir el programa anual de evaluaciones a proponer al primer ministro, registrar los reportes de evaluación de los comités y ofrecer el servicio de consulta metodológica por solicitud. El Comisariado del Plan actúa como apoyo secretarial del CNE y publica los informes de evaluación, junto con los comentarios de este Comité y de los ministros.

El procedimiento en marcha implica una toma de decisión respecto de lo que será evaluado (leyes, políticas, programas), seguida del nombramiento por el gobierno del presidente del dispositivo o comité de conducción de la

evaluación, una instancia independiente compuesta en cada caso por 12 a 30 miembros designados por el primer ministro a título personal (profesionales, diputados locales, expertos, servidores públicos, representantes de la sociedad civil y grupos de interés, entre otros) con objeto de monitorear la evaluación (definir metas, realizar la planeación de la investigación, efectuar audiencias, visitas de trabajo, etcétera), interpretar e integrar el conocimiento validado y escribir el informe final. Para el cumplimiento de su objetivo, el dispositivo de evaluación es responsable de escuchar a los diversos actores implicados, previo a la discusión de los problemas específicos que serán analizados en el trabajo y de la coordinación del estudio preliminar que permitirá la elaboración de los términos de referencia (TR). Posteriormente elige a los encargados de realizar la evaluación, que pueden ser órganos de la administración pública, consultorías privadas, equipos universitarios o de investigación.<sup>3</sup>

De esa forma puede decirse que en la evaluación participaban cuatro tipos actores (autoridades públicas, instancias de evaluación interna, comités de evaluación *ad hoc* y expertos profesionales), que cumplen tres roles bien diferenciados (una instancia que solicita la evaluación, otra que la coordina y un órgano científico o experto que la realiza), con mecanismos de coordinación

<sup>3</sup> Como ejemplo de la composición del dispositivo de evaluación, el correspondiente a la evaluación de la Ley Quinquenal del Trabajo, el Empleo y la Formación Profesional contó con seis miembros de la Administración Pública (Ministerio correspondiente y ministerios de Educación Nacional, Asuntos Sociales y Presupuesto, delegado interministerial y Comisariado General del Plan), y nueve personalidades calificadas (un experto, dos investigadores universitarios, dos diputados locales, un operador, un empresario y dos representantes de asociaciones). Esta instancia fue asistida por un redactor general y varios redactores auxiliares puestos a disposición por los ministerios involucrados (CIME-Primer Ministro-Comisariado General del Plan, 1997). Otro ejemplo relevante lo constituyó la evaluación de la política de lucha contra el síndrome de inmunodeficiencia adquirida (sida) 1994-2000. En este caso se buscó que la instancia no estuviera conformada por especialistas en la epidemia sino por investigadores o representantes de asociaciones. Se integró con participación de diferentes disciplinas de las ciencias sociales intervinientes en el campo: epidemiología, sociología, economía, derecho, medicina general, psiquiatría, medicina social y preventiva, además de los representantes de los administradores, miembros de salud, justicia y educación y el jefe de la Dirección de Sida. Su primer presidente renunció a los ocho meses por no poder garantizar la calidad científica del informe en el tiempo disponible (CNE-Comisariado General del Plan, 2002).

entre ellos. De esta organización depende en buena parte el éxito del trabajo, mismo que está, a su vez, marcado por las características de quien preside la instancia de evaluación *ad hoc*.

En 1999 se consideraba que la evaluación se había vuelto una realidad en el país, aunque Francia seguía presentando un retraso en materia de rendición de cuentas de la gestión pública a la sociedad. Thoenig (1998) había caracterizado el dispositivo de evaluación francés como extremadamente centralizado y dependiente de los decisores gubernamentales, confrontado a dudas sobre su legitimidad y utilidad al no poder garantizar la eficiencia y la eficacia de los actos públicos, razón por la que se debía desmitificarla al mismo tiempo que difundirla.

Freyssinet (1998) señalaba que, a pesar de que la evaluación es estructurada y financiada por el Estado, la paradoja es que sus resultados son muy críticos, al poner el acento en las desigualdades, disfuncionamientos e ineficacias. Piensa que esto se explica por la tradición de independencia y deontología de quienes componen las instancias de evaluación, entre cuyos miembros aparecen fuertes contradicciones.

También en 1999 hay que resaltar la puesta en marcha por el parlamento de una Misión de evaluación y control del gasto público (MCE) y la creación de la Sociedad Francesa de Evaluación (SFE), con la integración de 250 universitarios, consultores y miembros de la Administración Pública, que ya llegaban a 300 en 2002.

Al cierre del siglo, nadie discutía la importancia de la evaluación, pero su arraigo era meramente institucional; por otro lado, la experimentación había sido totalmente desechada y sustituida por el “pluralismo a la francesa”, también el papel que originalmente cumplieron los economistas como evaluadores, quienes fueron reemplazados poco a poco por los sociólogos. El fuerte peso relativo del Ejecutivo sobre el Legislativo, así como su centralización histórica, había dificultado un desarrollo que tampoco había contado con una importante demanda ciudadana.

En agosto de 2000 se produjo una circular del primer ministro sobre la evaluación de los contratos de Plan Estado-región que relanza los procedimientos impulsados por el Comisariado General del Plan. El número de evaluaciones realizadas en Francia desde 1980 alcanzaba en 2002 la cifra aproximada de 1 500, con un amplio espectro de métodos y en la mayoría de los campos, incluyendo políticas locales (42 en total), ministeriales e interministeriales.

Hespel sintetiza el desarrollo metodológico e institucional que la evaluación ha logrado en el último decenio del siglo XX, pero advierte que las aportaciones logradas “indican la necesidad de atemperar con pragmatismo algunas ambiciones iniciales” (Hespel, 2002:51, traducción libre). Y es que se han difundido métodos y criterios de evaluación, se han formalizado los procesos en todos los ministerios, se ha desarrollado una *expertise* universitaria y administrativa, la Sociedad Francesa de la Evaluación ha realizado exitosos coloquios, se incrementó sustantivamente el número de evaluaciones interministeriales (15 en tres años), y se ha fortalecido el papel del Parlamento en relación con las finanzas públicas a partir de la reforma de su ordenamiento orgánico en 2001; pero, la cultura de gestión pública asociada con la evaluación está desigualmente desarrollada. Se trata de un déficit de prácticas (no de procedimientos formales), lo que no impide que tanto la izquierda como la derecha promuevan su fortalecimiento. Ha servido al menos para romper las barreras entre las instituciones que necesitan compartir la evaluación.

Como contraparte, los recursos que el CNE y el Comisariado destinaban a la evaluación seguían siendo muy limitados (unas 20 y 15 personas respectivamente en 2002, a las que se sumaban aproximadamente 150 miembros de las instancias de evaluaciones en curso, sin remuneración), la *expertise* se había desarrollado pero alcanzaba apenas unas decenas o una centena de personas, los recursos financieros eran insuficientes, y el CNE no contaba con suficiente independencia del Ejecutivo, ya que dependía del Comisariado y necesitaba del Acuerdo del primer ministro para solicitar una evaluación.

Thoenig (2002) percibía un desencanto de quienes esperaban un pilotaje más científico de los asuntos públicos, tanto a nivel gubernamental como del debate político. Algunos de los problemas que observaba en el estancamiento de la evaluación se relacionaban con su institucionalización formal exagerada (procedimientos), los obstáculos interpuestos para impedirla por los juegos de poder, la existencia de funcionarios y consultores que denominaban evaluación a estudios que no lo eran, su déficit de profesionalismo, la falta de recursos para su desarrollo y la reducida utilización de sus resultados. Afirmaba que se habían multiplicado los excesos como la experimentación de la década de 1960 en Estados Unidos, la insistencia tatcheriana del *value for money* en Reino Unido y su institucionalización por decreto en Francia.

Sin embargo, esto no significaba para el autor que, en determinadas condiciones, las evaluaciones no contribuyeran a aportar “un poco de conocimiento razonado y riguroso” a los juicios que orientan la acción pública, tanto en la perspectiva de la gestión como en la dirección política del gobierno.

[En particular] el uso de la evaluación se aceleraría si existiera una comunidad profesional digna de ese nombre [y] uno o dos órganos instrumentalmente competentes instalados en el corazón de la maquinaria gubernamental y que gozaran de una especie de magistratura moral (Thoening, 2002: 458, traducción libre).

Thoening identificaba dos razones básicas para promover una mayor dosis de evaluación: nuevamente, el aprendizaje organizacional y de sus actores; y una mejor fundamentación de la autoridad política. La productividad y la eficiencia no entusiasman al ciudadano, tampoco éste valora de inmediato las ganancias en transparencia ni aprueba sus costos, lo que, en ausencia de sus referencias clásicas, ha restado fundamentación a la evaluación. Pero el Estado no sólo es cuestionado en términos de eficiencia, sino también en su capacidad para tratar los problemas públicos, lo que, para el autor, abre la segunda posibilidad para la evaluación: participar en la construcción de una nueva autoridad política basada en un gobierno compartido con la sociedad, facilitar la identificación de las partes que se deben involucrar, y desarrollar mesas de debate sobre los problemas.

Los cambios mencionados requerían transformaciones profundas en los servicios de *expertise* interna de los servicios centrales del Estado, cuya tendencia a la regulación tecnocrática y corporativista persistía desde la década de 1950 en que se impuso la planificación y la racionalización de las elecciones presupuestales. Este enfoque es aún liderado por pequeños grupos elitistas (École Polytechnique), cuyo monopolio del saber se les escapa cada vez más.

Thoening subrayaba también que los contratos no garantizaban el éxito y calidad de las evaluaciones, que los ministerios centrales enfatizaban la medición de resultados más que de impactos y tendían a considerar que la información más sofisticada era mejor, que la toma de decisiones seguía siendo política, por lo que la evaluación era sólo una fuente adicional de insumos que debía ayudar a los procesos de aprendizaje para pensar bien y ser oportunos. Consideraba que

la paradoja era que, a pesar de estar tan ampliamente difundida, la evaluación seguía siendo poco usada como herramienta gubernamental.

En los años siguientes a estas reflexiones, se han producido modificaciones trascendentes: en junio de 2004 se presentó un informe del Senado en el que se incluía la propuesta de creación de una oficina parlamentaria de evaluación de las políticas públicas y del control del gasto público, discusión a la que no se dio continuidad. En junio de 2005, la Fundación Concorde<sup>4</sup> organizó un coloquio en el Senado con la participación de renombrados políticos<sup>5</sup> en su favor. Recientemente se presentó una nueva propuesta (Comisión Attali) que después de fuertes discusiones<sup>6</sup> fue rechazada en mayo, pero con una fuerte toma de conciencia de la responsabilidad parlamentaria en la materia, que hasta el momento no había sido asumida.

En marzo de 2006, el Comisariado General del Plan fue sustituido por el Centro de Análisis Estratégico, con lo cual también desapareció el dispositivo central de evaluación que se concretaba en el Consejo Nacional de Evaluación ([www.plan.gouv.fr](http://www.plan.gouv.fr)), el que fue reemplazado por la Misión de Evaluación de las Políticas Públicas. Al mismo tiempo, el tema fue introducido en el texto constitucional (artículos 24, 47-2, 48 y 51-2). En consecuencia, el Parlamento se volvió responsable de controlar la acción del gobierno y evaluar las políticas públicas, con la asistencia de la Corte de Cuentas. Nuevamente se diseñaba un sistema de “evaluación a la francesa”, en lugar de reproducir el dispositivo del Reino Unido, dependiente exclusivamente del Parlamento.

Poco más de un año después, el presidente Nicolás Sarkozy elevó la importancia de la evaluación a rango ministerial al encargar la prospectiva, la evaluación de políticas públicas y el desarrollo de la economía numérica a Eric Besson, uno de los secretarios de Estado del primer ministro.

<sup>4</sup> Fundación Concorde es un “tanque pensante” compuesto por más de 2 500 empresarios, ministros, parlamentarios, representantes locales, etcétera, creado en 1997, con objeto de enriquecer el debate público y sugerir propuestas innovadoras a los decisores.

<sup>5</sup> Figuraron como participantes Michel Rocard (ex primer ministro y ex presidente del Consejo Económico y Social de la Unión Europea) y Raymond Barré (economista y ex primer ministro).

<sup>6</sup> Reservas de muchos parlamentarios frente al nuevo organismo y preferencia por mantener el control del gasto público en mano de la Corte de Cuentas.

En 2008, en el coloquio organizado por la SFE y la Sociedad Alemana de Evaluación<sup>7</sup> se percibía un crecimiento importante de las organizaciones y trabajos especializados en el tema, una tendencia a redescubrir métodos que prácticamente habían dejado de usarse y una relativa preocupación por el futuro de la evaluación en Francia.

El secretario Besson afirmó que, además del ministerio, existían otras 50 organizaciones francesas dedicadas a la evaluación. La Secretaría había publicado ya varias evaluaciones y tenía otras en proceso, relacionadas con el desarrollo profesional, el empleo o el precio de los alimentos. Para el futuro, se pensaba preparar un Plan Nacional de Evaluación, diseñar un sitio de internet donde se pudieran consultar herramientas e informes, e impulsar la formación en el tema. También se ampliaría el campo en el que se realizaban trabajos y se aumentaría su difusión pública. Era posible percibir un nuevo énfasis en la realización de evaluaciones *ex ante* y una insistencia de usar métodos multicriterios.

En relación con la evaluación de impacto revivían los métodos experimentales, a nivel universitario<sup>8</sup> y en la planeación gubernamental, mismos que habían sido desechados en Francia después de la década de 1970.

Ohana (2008) manifestaba los desiguales avances franceses por regiones y la existencia de un gasto público creciente y no controlado, los nuevos impactos no evaluados y una voluntad política de darle mayor seguimiento a estos asuntos. Ante esta situación, era partidario de crear la rechazada oficina parlamentaria de evaluación de las políticas públicas y del control del gasto público del tipo de la NAO del Reino Unido.

Si bien Ohana reconocía que el gobierno y la Corte de Cuentas están perfectamente capacitados para ejercer el control *ex post*, consideraba que el Parlamento necesita cubrir otras necesidades como considerar los impactos de sus decisiones, incrementar su transparencia ante la opinión pública, asegurar la eficacia y eficiencia del gasto público, entre otros. Para lograrlo propone compartir responsabilidades, clarificar conceptos sin tender a oponerlos (evaluación, control, auditoría, etcétera) o motivar al parlamento a utilizar estas herramientas al servicio de la democracia en los espacios en que puede actuar (Cardozo, 2009).

<sup>7</sup> Colloque Européen "L'évaluation des politiques publiques en Europe. Cultures et futurs", SFE-DeGEval, Estrasburgo, Francia, del 2 al 4 de julio de 2008.

<sup>8</sup> En combinación con la Universidad de Berkeley o el Massachusetts Institute of Technology de los Estados Unidos.

*España: ¿una formalidad para cumplir con la Unión Europea?*

La cultura de la evaluación ha estado poco desarrollada en España y se ha iniciado a raíz de su incorporación a la UE. El proceso de institucionalización de la misma comienza con el Real Decreto número 951 de 2005 sobre Mejora de la Calidad en la Administración del Estado y con los trabajos sobre el análisis del impacto normativo encomendados por Acuerdo del Consejo de Ministros, de febrero del mismo año. Hasta esa fecha, si bien existían instituciones que realizaban evaluaciones (Ministerio de Economía, Ministerio de las Administraciones Públicas), su dimensión era más bien sectorial o relacionada con las políticas de gasto. Recién en 2006, en el marco del Programa Nacional de Reformas, fue creada la Agencia Gubernamental para la Evaluación de Políticas Públicas y la Calidad de los Servicios (Aeval), como organismo público dependiente del Ministerio de las Administraciones Públicas, y posteriormente del Ministerio de política territorial y administración pública, es decir a nivel del Poder Ejecutivo (aún no hay avances en el Legislativo) pero sin dependencia directa respecto de ningún ministerio sustantivo. La misma se constituyó en enero de 2007.

Las principales funciones que se le asignaron fueron:

- Impulsar y fomentar la cultura de la evaluación de la gestión de la calidad.
- Proponer indicadores, sistemas de evaluación, metodologías y preparar guías de trabajo.
- Realizar la evaluación, anualmente decidida por el Consejo de Ministros, de las políticas y de los programas gubernamentales.
- Conducir una evaluación anual de los resultados del Programa Nacional de Reforma.
- Preparar y difundir un informe de evaluación del conjunto de servicios públicos analizados cada año por el Observatorio de la Calidad de los Servicios Públicos (Garde, 2004).

La Agencia concibió a la evaluación como una actividad institucional, enmarcada dentro de un proyecto político de gobernanza, modernización e innovación administrativa, que se propuso desarrollar con un enfoque pluralista y participativo de la ciudadanía y los distintos sectores sociales, de carácter estratégico y prospectivo. Asimismo, se definió por mantener una perspectiva

abierta a la colaboración con los órganos de la Administración general del Estado, las Comunidades autónomas y las Administraciones locales.

El Consejo de Ministros decidió que durante su primer año de funcionamiento debía realizar siete evaluaciones, entre las que se cubrían tres de las cuatro áreas definidas como prioritarias por la UE para la aplicación de la llamada Estrategia de Lisboa para el mejoramiento del bienestar social: energía y cambio climático; investigación y desarrollo; actividad empresarial y emprendedora (Cardozo, 2009). Éstas, y una octava que se agregó posteriormente, estaban en proceso al cumplirse su primer aniversario (Agencia de Evaluación y Calidad, 2008); sin embargo, el Informe de 2007 indicaba que todas se habían cumplido en los plazos previstos. En 2008 se reportaron 14 evaluaciones realizadas en los dos primeros años de actividad; a las que, en 2009, se agregan ocho más. En 2010 se llevaron a cabo otras seis. La mayoría de ellas ha estado relacionada con el Programa Nacional de Reformas y la Estrategia Española de Desarrollo Sostenible; todas habían sido desarrolladas por personal de la propia Agencia [<http://www.aeval.es>].

### **Evolución global europea**

En 2008, el Barómetro de la evaluación de políticas públicas en Europa (Ernst y Young, 2008), resaltó los avances realizados en países como Suecia o Alemania, y también en otros de más reciente inmersión en el tema como Polonia. Presentó también algunos resultados globales basados en trabajos de investigación, entrevistas a informantes-clave y encuesta en línea,<sup>9</sup> entre los que se destacan:

- “Las prácticas de la Comisión Europea constituyen un referente respetado por los países, y los programas europeos son los más evaluados.
- La identificación de tres grupos diferentes de países: *a*) los anglosajones con su enfoque de NPM y la eficiencia del *value for money*, *b*) los continentales

<sup>9</sup> Incluyó 1 050 decisores, funcionarios y universitarios de diez Estados miembros de la Unión Europea y un candidato, así como integrantes de la Comisión Europea; 44% de quienes respondieron son franceses e italianos. A esto se agregó información complementaria de Reino Unido y los países nórdicos.

orientados a una gestión por resultados, y *c*) los nuevos miembros, impulsados a raíz de su adhesión a la Unión Europea (UE).

- Una cultura y sensibilidad creciente ante la evaluación de políticas públicas (86% de las respuestas), debido a las exigencias impuestas por las autoridades europeas (56%).
- La creencia de que un buen evaluador debe tener sólidos conocimientos sectoriales y experiencia metodológica (42%), mientras que otros piensan que la última es suficiente (30%).
- La idea de que la evaluación sirve para mejorar la gestión, la eficacia y la eficiencia (65%).
- La opinión de que se trata de una práctica evaluativa madura (27%) o bien que todavía se encuentra en fase de desarrollo (58%).
- La exigencia sistemática de independencia de los evaluadores (51%), o sólo a menudo o a veces (42%).
- Los Poderes Legislativos utilizan poco a la evaluación para hacer contrapesos a los Ejecutivos (17%).
- Los informes de evaluación se consideran de alta calidad (75%), pero las recomendaciones sólo se toman en cuenta a veces (60%).
- Las principales técnicas de recolección de datos son: cuestionarios (64%) y entrevistas individuales (59%).
- Las principales herramientas de análisis son: estadística (49%), análisis multicriterio (37%), comparación con buenas prácticas (35%), análisis costo-beneficio y costo-eficacia (27%), análisis econométrico (12%).
- Las mayores dificultades metodológicas se relacionan con: evaluación de impactos socioeconómicos de largo plazo (46%), cuantificación de resultados, impactos y realizaciones (44%), e indicadores de logro mal definidos (44%).
- La difusión de resultados se realiza en forma sistemática (27%), a menudo (19%) o a veces (32%).

Finalmente, es importante destacar que, entre las preocupaciones muy presentes en el contexto europeo, y francés en especial, encontramos el tema de la profesionalización de los evaluadores y también de quienes coordinan (pilotean) sus trabajos desde las instituciones contratantes. Tourmen (2007) analiza diferentes situaciones de evaluación, los tipos de conocimientos y habilidades

necesarios para tener éxito en el trabajo, comparando personas con experiencia y debutantes (recursos técnicos, metodológicos, políticos y reflexivos), y la forma de desarrollar dichas competencias. Continuando en este sentido, Tourman y Monnier (2008) identifican dos campos conceptuales donde el desempeño de quienes tienen o no experiencia es muy diferente: *a*) todo lo relativo a la factibilidad de la evaluación, a lo posible, medible, evaluable; y *b*) lo vinculado con la legitimidad, confiabilidad, credibilidad, incluso utilidad. Consideran que la experiencia aporta capacidad de diagnóstico, de anticipación y adaptación.

Las ideas anteriores han llevado al Centro Europeo de *Expertise* y de Evaluación de Políticas Públicas (Euréval), a ofrecer diversos cursos de formación, y a la Universidad Lumière-Lyon2, conjuntamente con el Instituto de Estudios Políticos, a crear en 2004 una Maestría en Evaluación y Seguimiento de Políticas Públicas.

En España se ha contado con tres programas: el Master de Evaluación de Programas y Políticas del Centro de Estudios de Gestión de la Universidad Complutense de Madrid; el Master Iberoamericano de Evaluación de Políticas Públicas de la Universidad Internacional de Andalucía y la Sociedad Española de Evaluación, con apoyo de otras instituciones; y el Curso de Experto en Evaluación de Políticas Públicas y Programas Sociales de la Universidad de Almería, mismo que recientemente ha dejado de impartirse (Cardozo, 2009).

## **Comentarios finales**

El capítulo nos muestra la promoción que los organismos internacionales hacen del tema de la evaluación, a pesar de la variabilidad de sus enfoques. Por su parte, los países desarrollados revisados nos aportan aprendizajes de gran utilidad.

Estados Unidos nos ofrece información sobre los costos y peligros de una excesiva experimentación. Reino Unido destaca la utilidad de un trabajo de preparación de materiales y capacitación de funcionarios a todos los niveles, antes y durante la institucionalización de procesos de evaluación. Canadá, país en que ésta se desarrollaba como monopolio del Ejecutivo, nos advierte sobre el riesgo de transformar a la evaluación en un nuevo ritual que evite hacer críticas a los programas, no mida lo importante (la pertinencia de los programas y su impacto), no sugiera cambios sustantivos y no sea utilizada en los procesos

decisionarios. También sobre la importancia de realizar evaluaciones de políticas a nivel interministerial y de impulsar amplios procesos de participación social.

Francia nos enseña respecto de los peligros del centralismo y la falta de involucramiento del Poder Legislativo, el exceso de preocupación por los métodos de evaluación en detrimento de los problemas de organización y funcionamiento del contexto organizacional, la necesidad de contar con organismos, evaluadores y un fondo de financiamiento autónomos y de realizar evaluaciones selectivas de programas, entre otros.

España avanza en la realización de pocas evaluaciones anuales, cuyos informes y calidad se desconocen, aparentemente presionada por las normativas de la UE, como muchos de los países miembro.

Finalmente, Europa muestra importantes avances (aunque con distintos enfoques), incluso en los países que no pueden considerarse pioneros en el proceso (Polonia, Suecia, Alemania). Se percibe cultura y sensibilidad creciente, aunque se trata aún de un proceso en desarrollo. También dificultades con la evaluación de impacto y poca utilización de sus resultados por los poderes Legislativo y Ejecutivo, situación que veremos repetirse en las experiencias de otros contextos como las que prevalecen en América Latina, y México en particular, que serán objeto del siguiente capítulo.

## CAPÍTULO V

### Desarrollos empíricos de la evaluación en América Latina

Como ya se mencionó, la mayoría de los países ha intentado aplicar la evaluación a partir de las dos últimas décadas del siglo pasado. América Latina, región donde persisten resistentes problemas de pobreza, desempleo, bajos salarios, salud, educación, vivienda, etcétera, que sus gobiernos enfrentan con recursos muy limitados, cuenta con pocas experiencias de evaluación, y de limitado impacto.

México presenta múltiples intentos de evaluación, más o menos dispersos a lo largo del siglo XX, y un importante esfuerzo de sistematización de los procesos, mediante la modalidad de evaluación externa, a partir del año 2001.

#### **América Latina: crecimiento lento pero recientemente acelerado**

En América Latina, durante las décadas de 1950 y 1960 en que se produjeron reformas sociales populistas (con Perón en Argentina, Batlle en Uruguay, etcétera), la evaluación se aplicó en forma mínima. En la década de 1970, su uso se incrementó en algunos programas de desarrollo, promovido por organismos multilaterales y fundaciones internacionales que financiaban programas de apoyo.

En épocas más recientes, múltiples trabajos señalan el limitado avance de la evaluación en los países de la región. Como ya se mencionó, Sulbrandt (1993) consideraba que la mayoría de las políticas y los programas públicos de AL no se evaluaban, especialmente los programas sociales, y que, salvo excepciones, la evaluación no había conducido a provocar ningún cambio en el manejo de los programas ni en el aprendizaje de los ciudadanos.

Por su parte, Wiesner (2000) afirmaba que la evaluación por resultados no había logrado despegar debido a intereses particulares que conducían a la realización de evaluaciones espúreas, “precisas pero falsas” (pseudoevaluaciones) y débilmente vinculadas con los procesos de asignación de recursos, que impedían se convirtieran en un instrumento de reforma y modernización.

También Cohen (2001) manifestaba que no existía claridad sobre los beneficios provocados por la aplicación de programas para la reducción de la pobreza; lo que implicaba una crítica a la forma en que éstos se habían diseñado, pero también a la insuficiencia de las evaluaciones realizadas, especialmente en la medición de su impacto social.

Brown (2001) señalaba que en los países en desarrollo, AL y el Caribe en particular, los organismos ejecutores raras veces evaluaban los resultados de su gestión, siendo ésta promovida casi siempre por los organismos financieros (BM, BID, etcétera) para satisfacer sus propios requisitos de responsabilidad fiscal. Planteaba que la falta de evaluación de políticas y programas impedía una mayor utilidad de los presupuestos por programas y por resultados y provocaba importantes costos por falta de aprendizaje y repetición de errores.

De igual manera, Ospina (2001) consideraba que eran pocos los casos en AL (cita como ejemplos a Chile, Colombia y Costa Rica) en los que se podía hablar de una relativa evaluación sistemática en la administración pública y de la utilización de sus resultados, en sus cuatro niveles: políticas públicas, programas públicos, organizaciones públicas y desempeño de los empleados públicos. Pensaba que la evaluación de la gestión pública estaba dejando de ser un ejercicio técnico para transformarse en una herramienta política para resolver problemas de gobernabilidad y legitimidad, vinculada al enfoque de NPM. Así una mejor gestión del gasto público, afirmaba, mejoraría la imagen del Estado y aumentaría su legitimidad, liberaría recursos que podrían ser aplicados a satisfacer necesidades de grupos sociales de menores ingresos, aumentaba la equidad, etcétera. Sin embargo, no se reportaba ningún caso de AL en que estos resultados deseables se presentaran empíricamente. A pesar de la fortaleza de los esfuerzos chilenos y colombianos, Ospina dudaba aún de que se lograra promover la gestión orientada a resultados y el consecuente mejoramiento de la administración pública que constituía su objetivo.

Los factores que habrían desalentado a la evaluación eran: el proceso presupuestal, el alto costo de las evaluaciones, la toma de decisiones centralizada,

la falta de un interés continuado de parte de políticos y funcionarios públicos, la ausencia de un método eficaz para determinar el éxito o fracaso de los programas, las dificultades para utilizar indicadores directos, la insuficiencia de equipos técnicos, el interés exclusivo en el control de insumos y la observación de las normas, el privilegio del volumen sobre la calidad y de la cobertura sobre los resultados y la satisfacción, entre otros (Brown, 2001; Cohen, 2001).

Para Ospina el desafío era generar una dinámica que motivara a los líderes de una organización pública a demandar y usar información de manera explícita y sistemática para tomar decisiones gerenciales y políticas, y un diseño institucional que produjera los incentivos necesarios para que los usuarios vieran la importancia de acceder a la información, ya que, según Repetto (2003), tampoco la ciudadanía mostraba siempre el interés o las habilidades necesarias para utilizar plenamente la información generada en la evaluación.

En una publicación posterior (Cunill y Ospina, 2003), las autoras consideran que su estudio es el primero, hecho en profundidad, en la región, ya que insisten en que en AL se le ha prestado muy poca atención a la evaluación de la gestión pública, en comparación con la existente en Estados Unidos y los países integrantes de la Organización para la Cooperación y el Desarrollo Económicos.

En su trabajo analizan los tres casos de los países identificados por Ospina en el trabajo anterior y agregan el de Uruguay, en un intento por medir los esfuerzos de evaluación de la gestión pública en los procesos de modernización y democratización del Estado y la administración pública en la región, en los cuatro casos en que existen algunas iniciativas sólidas y herramientas de evaluación en operación.<sup>1</sup> Identifican las diferencias y semejanzas de los distintos sistemas diseñados y su impacto en el mejoramiento de dicha gestión, con la finalidad de difundir estas experiencias que pueden servir como oportunidad de aprendizaje para otros países. En el nivel meso, de evaluación de organizaciones y programas gubernamentales, destacan los avances del caso chileno.

Más recientemente (2008), las mismas autores realizaron un informe de avances que incluye los cuatro países anteriores y ocho adicionales. De sus dos trabajos extraemos los principales antecedentes de la evaluación de la gestión

<sup>1</sup> También reconocen avances menores en México, Brasil y Argentina, pero no los incluyen en el trabajo. Recientemente el CLAD ha agregado el estudio de nuevos casos nacionales (Argentina, Brasil y Honduras) que pueden consultarse en [www.clad.org.ve](http://www.clad.org.ve).

pública (que incluye a la evaluación de políticas y programas, pero también a la evaluación institucional y del desempeño individual) en once de los países analizados (ya que el duodécimo es México, que se presenta más adelante), mismos que se resumen en el Cuadro 1:

CUADRO 1  
*Evolución de la evaluación en América Latina*

Fecha	País	Actividad
1967-1968	Uruguay	Introduce el presupuesto por programas y la evaluación del cumplimiento de las metas y objetivos de todos los incisos del Presupuesto Nacional.
1974	Costa Rica	Encomienda al Ministerio de Planificación Nacional y Política Económica (Mideplan) la evaluación sistemática de los resultados de los planes, políticas y programas de desarrollo.
Década de 1980	Colombia	Se crea el Sistema de Evaluación y Seguimiento de Proyectos (Sisep), enfocado a los proyectos financiados por el BM y el BID, donde también participa el Banco de Proyectos de Inversión Nacional, encargado de las iniciativas del gobierno nacional y dirigido a la evaluación <i>ex-ante</i> .
1987	Uruguay	Retoma su proceso al aprobar la Ley de Rendición de Cuentas, que incluye información económico-financiera sobre el cumplimiento de metas y objetivos.
1990	Chile	La Oficina de Planificación Nacional se transforma en el Ministerio de Planificación y Cooperación (Mideplan), encargado de evaluar los proyectos de inversión financiados por el Estado.
1991	Colombia	La creación de un sistema de evaluación de la gestión y resultados de la administración pública, es encargada al Departamento Nacional de Planeación, como consecuencia de una reforma constitucional.
1992	Colombia	Se crea la División Especial de Evaluación y Gestión dentro del Departamento antes citado, institucionalizando las evaluaciones "ex-post".
1994	Colombia	Primera propuesta de Sistema Nacional de Evaluación de Resultados de la Gestión Pública (Sinergia), como herramienta para generar un cambio en la cultura de administración pública, abriendo la gestión a los ciudadanos, al ofrecerles amplia información ( <i>accountability</i> ). Este sistema careció del apoyo político necesario como consecuencia de la crisis de legitimidad que vivía el gobierno, pero fue validado por el BM y promovido como modelo ejemplar.

*continúa...*

CUADRO 1  
(continuación)

Fecha	País	Actividad
1994	Costa Rica	Su <i>Plan Nacional de Desarrollo 1994-1998</i> crea un sistema de evaluación de la gestión y un área específica de Evaluación y Seguimiento, que se acompaña de un decreto del Ejecutivo que da origen al Sistema Nacional de Evaluación (Sine). Este sistema surgió con un fuerte respaldo de la Presidencia de la República, mismo que ha mantenido a pesar de los cambios de administración.
1994	Chile	El gobierno inicia el impulso a la modernización de la gestión. El BM (2004) informa que se habría creado un sistema de Control de Gestión y Presupuestos por Resultados.
1994	Argentina	Se crea el Sistema de Seguimiento de la Ejecución Presupuestaria (SSEEP), regulado por la Ley de Administración Financiera y de los Sistemas de Control del Sector Público Nacional de 1992. Se basa en información trimestral de indicadores físicos de los programas.
1995	Chile	Se impulsa la utilización voluntaria de indicadores de desempeño, movimiento que inició desde 1991, con el regreso de la democracia.
1995- 2000	Uruguay	La evaluación recibió su mayor impulso a partir de la Ley de Presupuesto 1995-1999. En este último año se crea el Sistema Integrado de Información Financiera a cargo de la Oficina de Planeamiento y Presupuesto (OPP).
1997	Uruguay	Se crea el Sistema Informático de Evaluación de Gestión Pública (SEV), también centrado en la OPP.
1998	Uruguay	Se aprueba la Ley de Rendición de Cuentas.
1999	Uruguay	El SEV comienza sus operaciones, pero, a partir de estas fechas, el sistema tendió a declinar hasta 2005.
1997-1999	Chile	Se produce el primer intento por articular sus instrumentos de evaluación (de metas ministeriales, desempeño de funcionarios, etcétera) en un sistema integrado, con la creación del Sistema de Evaluación de Programas Gubernamentales y la implantación de los Balances de Gestión Integral. Los indicadores de desempeño se vuelven obligatorios, pero su uso es interrumpido dos años después.
2000	Colombia	Sinergia logra el apoyo político suficiente para pasar a ser orgánicamente una Dirección de Evaluación de Gestión y Resultados.

*continúa...*

CUADRO 1  
(continuación)

Fecha	País	Actividad
2000	Costa Rica	La nueva Constitución incorpora la evaluación por resultados y la rendición de cuentas como principios inspiradores de la institucionalidad democrática.
2000	Chile	El uso de los indicadores de desempeño se vuelve a adoptar en forma obligatoria para todas las entidades públicas, después de ajustarlos metodológicamente. Conjuntamente se adoptan los mecanismos de evaluación restantes. Sin embargo, todos estos mecanismos no integran un sistema único y articulado, distribuyéndose su responsabilidad entre varias entidades del gobierno central (Dirección de Presupuesto del Ministerio de Hacienda, División de Control de Gestión, Secretaría General de la Presidencia, Ministerio de Planificación, etcétera). Su apoyo legal está basado en leyes, decretos y acuerdos entre los poderes Ejecutivo y Legislativo. A partir de este año, el sistema tiende a integrarse en uno sólo, el Sistema de Control de Gestión (SCG).
2001	Nicaragua	Se concibe el Sistema Nacional de Seguimiento a Indicadores de Desarrollo (Sinasid) para la evaluación del <i>Plan Nacional de Desarrollo</i> y la política de combate a la pobreza; sufrirá diversas revisiones en 2003, 2005 y 2007.
2001	Honduras	Se crea el Sistema de Información de la Estrategia para la Reducción de la Pobreza (Sierp).
2002	Argentina	Se instala el Consejo Nacional de Coordinación de Políticas Sociales, organismo de la Presidencia de la Nación, con la finalidad de lograr una correcta y eficaz administración de los recursos del Estado destinados a la política social. Actividad desarrollada previamente por el Sistema de Información, Monitoreo y Evaluación de Programas sociales (Siempro).
2003	Brasil	El presidente Lula establece el Sistema de Seguimiento de Metas Presidenciales (SMMP).
2004	Perú	A partir de diversos instrumentos existentes desde 1997, se instala el Sistema de Seguimiento y Evaluación del Gasto público (SSEGP), con escasos avances hasta 2007, en que es renovado.
2004	Brasil	El Plan Plurianual (PPA) se integra con una Comisión de Seguimiento y Evaluación (CMA, por sus siglas en portugués) y de Unidades de Seguimiento y Evaluación (UMA) en cada ministerio sectorial, componiendo un sistema.

*continúa...*

CUADRO 1  
(termina)

Fecha	País	Actividad
2005	Uruguay	Con la llegada al gobierno del Frente Amplio se replantea la evaluación de la gestión, y se crea el Área de Gestión y Evaluación (AGEV), también a cargo de la OPP.
2006	Honduras	Se agrega el Sistema de Gerencia Basado en Gestión por Resultados (SGPR).
2006	Chile	El nuevo gobierno instala el Sistema de Seguimiento de la Programación Gubernamental (SPG).
2007	Brasil	Se comienza a abandonar el sistema de metas presidenciales por el de metas del Plan de Aceleración del Crecimiento (PAC).
2007	Paraguay	Se establece un Módulo de Control y Evaluación Presupuestaria (MCE) como parte del Sistema Integrado de Programación Presupuestaria.
2008	Bolivia	Se comienza el diseño de un Sistema Nacional de Monitoreo y Evaluación (SE) del nuevo <i>Plan Nacional de Desarrollo</i> (PND).

Fuente: elaboración propia con base en Cunill y Ospina (2003 y 2008).

A inicios del nuevo siglo, la situación en los cuatro países pioneros era la siguiente: el sistema colombiano abarcaba los niveles macro (políticas) y meso (unidades ejecutoras) de la gestión pública, aunque fue concebido para el primero, en el que no había logrado incidir para reorientar las políticas públicas. En su funcionamiento existían superposiciones de atribuciones, particularmente entre la Contraloría, el Ministerio de Hacienda y el equipo de Sinergia, quienes, en 2001, firmaron un convenio para resolver estos problemas. El sistema mantenía escasa vinculación con los poderes Legislativo y Judicial y se tenían planes para su descentralización a mediano plazo. Gozaba de un alto nivel de legitimidad porque se reconocía la necesidad de evaluar los resultados, aunque era poco conocido entre políticos, académicos y población en general.

En Costa Rica el diseño pretendía evaluar la gestión pública en sus tres niveles (macro, meso y micro) para ofrecer información a la población que garantizara la transparencia y legitimidad del sistema, así como el fortalecimiento democrático; pero hasta ese momento sólo se había evaluado parcialmente el nivel meso. Por

otra parte, no existían mecanismos de coordinación formal entre los diversos organismos de evaluación y control.

Mientras tanto, en Chile, la evaluación se hacía efectiva en los tres niveles de la gestión y contaba con amplia legitimidad, pero tampoco existía una articulación formal entre los distintos mecanismos, los que se había intentado complementar con la modernización de los servidores públicos, con un avance gradual de su descentralización y desconcentración a nivel regional y local, y la intención de incorporar a la ciudadanía al diseño y evaluación de las políticas y programas públicos. También se impulsó la adopción de un Sistema de Información para la Gestión Financiera del Estado (SIGFE).

Finalmente, Uruguay concentraba su evaluación en el nivel meso (aunque algunas entidades quedaban fuera del sistema), sin pretender evaluar las políticas públicas. El objetivo del Sistema de Evaluación de la Gestión Pública por Resultados (SEGPR) era proporcionar una herramienta para la toma de decisiones, que permitiera mejorar la asignación del gasto público y focalizar la acción en el ciudadano. Si bien existían cuatro organismos designados para participar en procesos de evaluación, no había habido problemas de superposición gracias a que, en 2001, dos de ellos aún no funcionaban y además sus atribuciones estaban claramente delimitadas por la ley. También se había intentado, aunque sin resultados positivos, incorporar a las ONG, que recibían subsidios públicos y contaban con sus propios mecanismos, al sistema de evaluación.

La comparación de los cuatro casos nacionales pioneros muestra que sus reformas han estado fuertemente influidas por el NPM; sin embargo, no son uniformes en diversos aspectos:

- Salvo el caso de Colombia que resolvió diseñar su sistema en forma independiente de toda experiencia previa, los demás países tomaron éstas en cuenta, particularmente la de Gran Bretaña en los casos de Costa Rica y Chile. El último, además, intercambió profesionales con Australia y Nueva Zelanda; en Uruguay también se sintió la influencia del NPM. Para Costa Rica fue relevante la experiencia de Colombia, y para Uruguay, la de Chile.
- Hay un grupo de países que tomó al proceso presupuestario como eje básico de la misma (modelo Presupuesto, en términos de Cunill y Ospina), modificando su énfasis en el control de los gastos por una mayor preocupación de la evaluación de los resultados medibles (Uruguay y Chile); mientras los

dos restantes buscaron la creación de mecanismos que aumenten la capacidad institucional del Estado mediante la modernización de las organizaciones públicas, que refuercen sus decisiones políticas, más que las presupuestarias (modelo Plan, en la misma fuente mencionada).

- En Uruguay, Colombia y Costa Rica se crea un sistema nacional de evaluación, centralizado y dirigido desde un nivel estratégico del aparato estatal sobre el conjunto de la administración pública del país; en Chile se trata de un conjunto de continuas iniciativas relativamente independientes, sin la pretensión anterior.
- Los cuatro sistemas iniciales presentados fueron introducidos en forma gradual, mediante pruebas piloto, salvo en el caso uruguayo que se implantó simultáneamente, en un tiempo muy reducido, en todas las unidades Ejecutoras del Presupuesto Nacional, lo que dificultó la acumulación de aprendizajes.
- Su incorporación fue acompañada en todos los casos de procesos de capacitación, realización de seminarios, talleres, difusión de otras experiencias, distribución de manuales, instructivos, guías metodológicas, libros, videos, etcétera.
- Costa Rica y Uruguay recibieron apoyo técnico del BID, Colombia del BM y Chile de ambos.
- En los países analizados, el sistema cubría sólo el nivel central de gobierno, salvo Chile, que mostraba algunos avances adicionales a nivel regional y Costa Rica que incluía algunas instituciones descentralizadas. Se evaluaban organizaciones y programas seleccionados, mediante consultoría externa y algunos monitoreos a cargo de los ministerios respectivos.
- Mientras Uruguay había definido un gran número de indicadores que se consideran poco efectivos, Chile y Colombia habían aprendido a limitar su número y mejorar su diseño, aumentar la calidad de la información y facilitar su utilización.
- La legitimidad en el diseño de los anteriores dependía de la participación ciudadana en esta etapa, asunto que Costa Rica comenzaba a considerar. Su debate requeriría de una evaluación de tipo pluralista como la propuesta por Monnier.
- El mecanismo empleado para informar resultados era principalmente la presentación de informes al Congreso. En el caso colombiano se realizaba

anualmente e incluía un resumen de todas las evaluaciones realizadas, el análisis del grado de cumplimiento del plan y propuestas de ajustes. En Chile, los presentaba cada institución, utilizando como instrumento el Balance de Gestión Integral, que sintetizaba los resultados de las evaluaciones. Uruguay presentaba similar información en el *Documento de rendición de cuentas*. Finalmente, Colombia entregaba, cada cuatro meses, al Consejo Nacional de Política Económica y Social, información del seguimiento realizado, y anualmente, de la evaluación efectuada.

- Como incentivo a las buenas prácticas, los cuatro países utilizaban la difusión pública de los resultados de las evaluaciones de desempeño institucional. Las vías de divulgación generalizada eran las páginas web institucionales; Costa Rica utilizaba también conferencias de prensa sobre el tema. En Uruguay, la principal forma de acceso a esta información era la prensa, pero ésta le había prestado muy poca atención al tema. Esta limitación se agudizaba debido al lenguaje excesivamente técnico en que se redactaban los documentos de difusión. En el esquema chileno se vinculaban los logros alcanzados con incentivos de diversos tipos, como los Programas de mejoramiento de gestión (PMG), que relacionan el nivel de desempeño con las decisiones sobre incrementos salariales y los resultados de las evaluaciones de programas gubernamentales, que se utilizaban para definir su marco presupuestario, a partir de un sistema de “premios” y “castigos”. Su utilización podía convertirla en una actividad amenazante para la gerencia, que generaba respuesta defensivas, en lugar de coadyuvar a mejorar el rendimiento, por lo que parecía más efectivo el uso de incentivos simbólicos de los otros países.
- Si no se hacía uso de la información con algún tipo de consecuencias directas, como había sucedido en los casos de Colombia y Uruguay cuando el resultado no era el esperado, y no se realizaba un aprendizaje ciudadano de los instrumentos de evaluación, el sistema no se tomaba con seriedad. En los dos países citados y Chile, aunque la información se publicara en Internet, no había evidencia de que los ciudadanos accedieran a ella, lo que Costa Rica había resuelto, en parte, con sus publicaciones en la prensa.
- En los últimos años del siglo XX se habían fortalecido las organizaciones de la sociedad civil especializadas en la evaluación de la gestión pública, lo que no sucedía en el caso de los legisladores. Ellos parecían ser actores privilegiados sólo bajo el modelo Presupuesto, sin embargo, enfrentaban fuertes limitaciones

tanto en Chile como en Uruguay para afectar la asignación de recursos; mientras que en el modelo Plan ese lugar lo ocupaba la Presidencia.

- En cuanto a los logros reales de los sistemas de evaluación descriptos, aún quedaba mucho por hacer para que se acercaran a lo esperado, sin embargo, recién estaban en proceso de implementación:
  - Colombia había avanzado en la generación de una cultura orientada a resultados y rendición de cuentas y logró que los valores en que se sustentaba la evaluación cambiaran de la eficiencia del gasto a los resultados e impactos. También presentaba un desarrollo importante de los instrumentos, una rápida consolidación del sistema y una definición más clara de responsabilidades; pero, las conclusiones de las evaluaciones rara vez eran incorporadas en los procesos decisorios.
  - En Costa Rica el sistema había contribuido a promover la conducción estratégica del gobierno y la rendición de cuentas, pero poco a desarrollar capacidades en materia de programación presupuestaria, otro de sus objetivos.
  - El incremento de la eficiencia de la gestión pública, la medición de sus resultados y el aumento de la transparencia, constituían los principales logros chilenos, a pesar de su falta de integración. Durante el año 2002, se habrían evaluado doce de sus programas, ocho de los cuales perseguían objetivos sociales.
  - En el caso uruguayo, estaba funcionando un sistema de indicadores de desempeño, pero no se había alcanzado el nivel de una verdadera gestión del desempeño, donde la evaluación se transformara en aprendizaje y se utilizara en la toma de decisiones, principalmente a nivel estratégico. No existía ningún mecanismo destinado a incentivar su uso. Tampoco su Parlamento contaba con información suficiente para analizar el cumplimiento de metas.

Contrario a lo esperado por Cunill y Ospina (2003), el sistema nacional y coherente de Uruguay parece haber sido menos efectivo que las diversas iniciativas inconexas de Chile. En realidad, las autoras mencionadas reconocían que no contaban aún con evidencia suficiente en ninguno de los países que instalaron sistemas de evaluación para asegurar que hubieran sido afectadas positivamente la calidad de los servicios o la productividad del sector público.

Entre los desafíos comunes a estos países figuraban el mejoramiento de los sistemas de información, así como su calidad y confiabilidad, las estrategias de difusión de resultados y el desarrollo o consolidación de incentivos vinculados con el uso de la información generada sobre los resultados alcanzados.

A pesar de ello, Cunill y Ospina (2003) concluyeron que existían motivos para ser optimistas, ya que había suficiente consenso entre los funcionarios y gerentes públicos sobre la importancia de orientar su gestión a resultados, lo que constituía el primer paso hacia una cultura de nueva gestión pública y se percibía una notoria diferencia respecto de la situación prevaleciente cinco años atrás. Después del 2003, estas autoras han continuado el estudio de los procesos de evaluación que se desarrollan en la región y, como consecuencia de ello, pueden encontrarse documentos referidos a nuevos países como Argentina, Brasil, Perú, Bolivia, Paraguay, México, Honduras o Nicaragua [www.clad.org.ve]. Por ejemplo, en Brasil, la mayor parte de los programas de su Plan Plurianual se evalúan anualmente y su informe es entregado al Congreso para considerarlo en el ciclo presupuestal; sin embargo, tampoco hay evidencias de que sean tomados en cuenta.

Cinco años más tarde, las autoras (Cunill y Ospina, 2008), identificaron un aumento importante de los sistemas de evaluación en América Latina (Cuadro 1) y los clasificaron en función de su orientación como:

- Predominantemente política, en los casos de Colombia (Sinergia), Costa Rica (Sine), Nicaragua (Sinamid), Uruguay (SEV), Honduras (Sierp), Chile (SPG) y Brasil (SMMP).
- Predominantemente económica, en Chile (SCG), Perú (SSEGP), Argentina (SSEEP), y Paraguay (MCE).
- Expresamente económica y política, en Brasil (PPA) y Bolivia (SE).
- Social, en Chile (Mideplan), Brasil (SSEEP) y Argentina (Siempro).

Complementariamente, el BM (2004) señalaba que en Chile, al igual que en Estados Unidos, se había implementado un esquema gubernamental integral de M&E con base en la Administración Basada en Resultados y se habían capacitado 180 profesionales de diferentes instituciones y ministerios en materia de evaluación de programas. A mediados de 2004, dicho país puso también en marcha, con asistencia del BID, un sistema que incluye la selección de los gerentes

públicos por un Consejo de Alta Dirección Pública, cuyos integrantes son nombrados por el Senado, y que, en materia de evaluación, prevé su realización por una entidad independiente de los poderes Ejecutivo y Legislativo, similar al Rekenkamer en Holanda. Waissbluth (2004) consideró que se trataba de la reforma más amplia realizada en AL en las últimas décadas.

Por su parte, la experiencia argentina, iniciada en 2003 por la Jefatura de Gabinete de Ministros con apoyo del BM, es relatada por Mosqueira (2007), al igual que los casos de Chile y Colombia que califica de exitosos. En su trabajo revisa las condiciones argentinas que permitirían implantar exitosamente un modelo de Evaluación y Monitoreo (credibilidad y confianza, utilización de resultados en la toma de decisiones, existencia de incentivos y comportamientos adecuados, operación de las decisiones presupuestarias y la gestión administrativa, etcétera), así como las disfunciones existentes en la operación de su gestión pública, especialmente en su proceso presupuestario. Plantea que tanto el gobierno como la ciudadanía han empezado a entender que las evaluaciones pueden mejorar la relación Estado-ciudadano, por lo que se han diseñado programas como “Carta compromiso con el ciudadano” o “Gestión de resultados” y se han emitido decretos para aumentar la transparencia del gobierno (Cardozo, 2009).

Un elemento a resaltar en América Latina ha sido la creación en 2003 de una Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLAC), constituida en el marco de la Asamblea Inaugural de la Organización Internacional para la Cooperación en Evaluación (IOCE por sus siglas en inglés), con la participación de redes profesionales de Brasil, Costa Rica, Colombia y Perú. En 2004 llevó a cabo su Primera Conferencia Bienal con la participación de 24 países; la segunda, en 2007, fue organizada en Colombia, conjuntamente con la Asociación Internacional de Evaluación para el Desarrollo (IDEAS), y permitió un profundo diálogo entre evaluadores, formuladores y decidores de políticas. En ese momento ya contaba con la adhesión de 14 redes nacionales<sup>2</sup> (Cardozo, 2009). La tercera Conferencia tuvo lugar en Costa Rica en 2010.

<sup>2</sup> Las mismas provienen de Argentina, Bolivia, Brasil, Colombia, Chile, Ecuador, Venezuela, Uruguay, Perú, Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua. A éstas se agregó México como participante en la Conferencia Bienal.

ReLAC constituía el único espacio relevante de encuentro de experiencias a nivel latinoamericano, a la que se ha sumado recientemente la Red Latinoamericana y del Caribe de Monitoreo y Evaluación (Redlacme).

### **México: una sistematización largamente postergada**

Desde el siglo pasado muchos programas públicos, y especialmente sociales, han sido desarrollados por diversas secretarías de Estado (Salud, Educación Pública, Desarrollo Social, etcétera), organismos descentralizados (Instituto Mexicano del Seguro Social, Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, Instituto Nacional Indigenista, etcétera) y empresas de participación estatal mayoritaria (Comisión Nacional de Subsistencias Populares, Conasupo; Leche Industrializada Conasupo, Liconsa; Distribuidora Conasupo, Diconsa, etcétera), en México.

A partir de 1917 se realizaron intentos por crear un organismo que midiera, siquiera parcialmente, los resultados de la gestión gubernamental, poniendo énfasis en el manejo de los recursos. Sin embargo, la lógica sexenal hizo que cada administración llegara con su nuevo personal público y reorientara los programas gubernamentales para ponerles su sello propio, provocando el desinterés por evaluar lo hecho por la anterior, situación que prevaleció hasta iniciado el siglo XXI.

De esta forma, lo realizado en materia de evaluación fue poco, parcial y estuvo a cargo de diversas instancias como la Secretaría del Patrimonio Nacional, la Secretaría de la Presidencia o la Secretaría de Hacienda y Crédito Público (SHCP); desconociéndose realmente los niveles de eficiencia y de impacto de los programas gubernamentales en los problemas que pretendían resolver.

Los intentos de reformas, especialmente dirigidos a los procesos de planeación de los que la evaluación formó parte, comienzan en la administración de López Portillo, y su origen podría ubicarse en los trabajos realizados, desde 1965, por la Comisión de Administración Pública (CAP), a la que pertenecieron tanto él como varios colaboradores suyos.

En el Cuadro 2 se resumen los principales antecedentes vinculados con el tema de la evaluación de la gestión gubernamental en general y de los programas de desarrollo social en particular.<sup>3</sup>

CUADRO 2  
*Antecedentes de la evaluación en México*

---

HACIA FINALES DE LA DÉCADA DE 1960

*Evaluaciones independientes:* las primeras evaluaciones fueron desarrolladas en forma independiente por universidades y centros de investigación y continuaron realizándose hasta la última década del siglo XX, en pocas ocasiones y respondiendo a iniciativas de los propios investigadores. Así encontramos el primer antecedente, en que la Universidad Nacional Autónoma de México (UNAM) realizó los esfuerzos dirigidos a la implantación de un Presupuesto por Programas.

1974, 1977

*Congreso de la Unión:* debido a las modificaciones introducidas en los años señalados, el Congreso de la Unión adquirió algunas facultades respecto a los organismos paraestatales, además del control general de sus ingresos y egresos, pudiendo citar a los directores y administradores de los mismos para que informaran cuando se discutiera una ley o se estudiara un negocio concerniente a su ramo y también para que, cuando lo pidiera una cuarta parte de los miembros de la Cámara de Diputados y la mitad de la de senadores, se integraran comisiones para investigar el funcionamiento de dichas entidades, y para poner en conocimiento del Ejecutivo el resultado de sus investigaciones.

1976-1982

*Administración pública:* el trabajo efectuado en la UNAM fue continuado, durante el sexenio de José López Portillo, quien creó la Secretaría de Programación y Presupuesto (SPP), encargada de la planeación, programación, presupuestación, control y evaluación del gasto público. La SPP contó con una Subsecretaría de Evaluación de los resultados de los programas gubernamentales, integrada sólo por una Dirección General, que sin mucha claridad por la falta de antecedentes, pretendió ir más allá de la vigilancia del gasto, identificando y corrigiendo errores, replanteando estrategias, etcétera, para lo que debía diseñar e implantar un sistema nacional de evaluación; sin embargo, sus funciones se duplicaron con las de la Coordinación de Evaluación, dependiente de la Secretaría de Presidencia, creada en 1977, y sus esfuerzos se diluyeron. La Subsecretaría de Evaluación acabó realizando estudios parciales y de poco impacto en el proceso.

*continúa...*

<sup>3</sup> No se incluyen los avances específicos en los sectores de salud y educación.

CUADRO 2  
(continuación)

---

1978

*Congreso de la Unión:* se fortaleció al Poder Legislativo en su control de las acciones del Ejecutivo, al aprobarse la Ley Orgánica de la Contaduría Mayor de Hacienda (CMH), órgano cuyos antecedentes pueden rastrearse hasta la administración colonial del siglo XV. La CMH fue el órgano técnico de la Cámara de Diputados (estuvo vigente hasta 1999), para el cumplimiento de sus facultades exclusivas como revisor del manejo de fondos de la Cuenta Pública del Gobierno Federal. Debía rendir un informe previo a la Comisión de Presupuesto y Cuenta Pública de la Cámara dentro de los 10 primeros días de noviembre, siguientes a la presentación de la Cuenta Pública y un informe definitivo a la misma Cámara sobre el resultado de la revisión, en los 10 primeros días de junio del año siguiente a su recepción (artículo 74 constitucional). En caso de presentarse problemas se determinaban y fincaban las responsabilidades correspondientes.

1980

*Participación ciudadana:* el gabinete instituye el mecanismo de participación y consulta a la ciudadanía, que tíbiamente había empezado a impulsar en la llamada “reforma de ventanillas”, cuyo único logro fue la simplificación de algunos trámites. Sin embargo, no se avanzó más allá de la definición de sus objetivos.

1980

*Administración pública:* se publica el primer *Plan global de desarrollo* y Miguel de la Madrid, al frente de la SPP, transforma las funciones de la Subsecretaría de Evaluación, responsabilizándola de preparar la información para la elaboración de la Cuenta Pública.

1982

*Administración pública:* el énfasis en la planeación del sexenio anterior no se abandona pero, la evaluación adquiere una centralidad que no había gozado nunca, para responder a la renovación moral ciudadana, cuando el presidente Miguel de la Madrid (1982-1988) crea la Secretaría General de la Federación (Secogef), con carácter de globalizadora y atribuciones normativas para el resto de los organismos gubernamentales, encargada de ejercer la fiscalización, el control interno global y la evaluación sobre la Administración Pública, mediante las contralorías internas de las dependencias

---

*continúa...*

CUADRO 2  
(continuación)

---

y entidades. Dicha Secretaría debía cumplir sus funciones al interior de la Administración Pública sin entrar en conflicto con la CMH, que lo hacía desde otro Poder, el Legislativo. También para evitar éstos, se disolvió la Coordinación de Evaluación en la Presidencia.

1983

*Administración pública:* frente a la severa crisis económica prevaleciente, la SPP pierde su énfasis en la planeación para pasar a encargarse de los programas de estabilización, saneamiento de las finanzas públicas, contención del gasto, etcétera. En este nuevo encuadre las funciones de evaluación se trasladaron completamente a la Secogef; sólo permaneció en la SPP la contabilidad gubernamental y el control sobre las adquisiciones, para cumplir con la elaboración de la Cuenta Pública. En la propia Secogef también se vivió el desvío de funciones de sus comisarios y delegados, que acabaron trabajando en la liquidación o fusión de empresas o en los recortes de personal, muy lejos de combatir la corrupción.

1984

*Administración pública y consulta ciudadana:* mediante la modificación del artículo 26 constitucional, fue puesto en marcha el Sistema Nacional de Planeación Democrática, a partir de las orientaciones de la Ley de Planeación y el *Plan Nacional de Desarrollo*, de 1983. Trataba de incorporar el carácter público y consensual del mismo, así como la participación de las entidades federativas en la planeación del desarrollo. Indicaba el mecanismo de consulta popular para recoger la participación social en las etapas de formulación, control y evaluación, el régimen de obligatoriedad de la planeación dentro del Sector Público Federal, el de concertación con otros niveles de gobierno y el inductivo para el ámbito no gubernamental.

1985

*Administración pública:* se realizó la primera y única reunión de evaluación del programa de simplificación administrativa de trámites, mismo que buscaba elevar la productividad y eficiencia del aparato estatal. Se trató de una mera descripción de avances.

---

*continúa...*

CUADRO 2  
(continuación)

---

1985

*Auditorías externas:* se amplía la utilización del mecanismo de contratación de despachos privados, que ya se había empleado en el sexenio previo, para la realización de auditorías sobre aspectos específicos de la gestión pública.

1986

*Asesoría sobre evaluación:* Diconsa solicitó al Departamento de Administración Pública del Centro de Investigación y Docencia Económicas (CIDE) una propuesta para evaluar su impacto socioeconómico y político. La misma fue presentada pero se desconoce si llegó a aplicarse.

1988

*Administración pública:* al finalizar el sexenio, la Secogef había dedicado mucho tiempo a la coordinación; su trabajo se había centrado más en la sanción de irregularidades que en su prevención; muchas áreas de evaluación habían desaparecido con la crisis económica; y, sólo en los últimos meses, se realizaron reuniones de autoevaluación de las entidades y dependencias gubernamentales, en que éstas se limitaron a presentar sus acciones en forma positiva y sin la menor autocrítica.

1988

*Administración pública:* el nuevo presidente Carlos Salinas de Gortari (1988-1994) pone en marcha el Programa Nacional de Solidaridad (Pronasol), manejado desde la SPP, a pesar de que ya no se trataba de una estrategia administrativa, sino de la forma en que se busca resolver el mayor problema social: la pobreza.

1989

*Administración pública:* se creó el Programa General de Simplificación de la Administración Pública Federal, que recogía algunos planteamientos del NPM e incluía entre sus objetivos: consolidar una cultura administrativa que propiciara la eficiencia y la productividad en la gestión de las instituciones gubernamentales, promover la evaluación de la calidad y la transparencia en la prestación de los servicios públicos y prevenir y combatir la corrupción. La responsable de aplicarlo era la entonces Secogef.

---

*continúa...*

CUADRO 2  
(continuación)

---

1990

*Administración pública:* se crea la Comisión Nacional de Derechos Humanos (CNDH), como órgano desconcentrado de la Secretaría de Gobernación, tomando como modelo el *ombudsman* escandinavo o el “defensor del pueblo” español, “como instrumento de la sociedad para vigilar y auxiliar a las autoridades en la observancia de la ley y en la protección de las garantías individuales” (Aguilar, 2000d:170-171).

1990

*Tesis sobre medición de la pobreza:* el doctor Miguel Székely Pardo (quien se desempeñó posteriormente como Subsecretario de Planeación, Prospectiva y Evaluación de la Sedesol), presenta su tesis “Identificación de la pobreza global y rural: un primer paso para incrementar el grado de acierto de los programas públicos” para obtener el grado de Maestro en Políticas Públicas por el Instituto Tecnológico Autónomo de México (ITAM). El mismo contiene como estudio de caso un acercamiento hacia la evaluación de los programas a cargo de Liconsa. Concluye en la necesidad de focalizar los apoyos hacia la población rural pobre.

1991

*Contraloría Social:* en ese mismo sexenio afloran los primeros atisbos de una contraloría social en México, mismos que inician al interior del Pronasol, asumiendo el principio de corresponsabilidad, como Vocales de Control y Vigilancia, elegidos por los miembros de los Comités de Solidaridad, a su vez, elegidos por las comunidades beneficiarias del programa. Se trata de órganos transitorios, limitados al control *ex-post* de los programas sociales, sin respaldo legal, personalidad jurídica ni especificación taxativa de atribuciones, por lo que obviamente no podían aplicar sanciones en forma directa, sino sólo informar a la Secogef de las irregularidades que identificaran. También había la confusión de que fungían como delegados del Sistema Integral de Control y Evaluación de la Gestión Pública y del Sistema Nacional de Quejas y Atención a la Ciudadanía, a cargo de la Secogef (Cunill, 2000).

---

*continúa...*

CUADRO 2  
(continuación)

---

1992

*Administración pública:* se disuelve la SPP y sus funciones son transferidas a la SHCP; también se crea la Secretaría de Desarrollo Social (Sedesol). Después de modificarse el artículo 102 de la Constitución, se otorga el carácter de organismo descentralizado con personalidad jurídica y patrimonio propio a la CNDH. Este último constituye uno de los organismos que ayudan al equilibrio de poderes y sus contrapesos, creado para la evaluación de un aspecto concreto.

1993

*Evaluación externa:* la Organización Internacional del Trabajo (OIT) encarga al doctor Miguel Székely Pardo la realización de una evaluación al Pronasol. El autor publica un documento de trabajo *El Programa Nacional de Solidaridad en México* que incluye un capítulo titulado “Hacia una evaluación del Programa Nacional de Solidaridad”, en que reconoce la inexistencia de elementos en la literatura disponible para responder a las preguntas que exceden a la información sobre el gasto ejercido y realiza un ensayo sobre los diferentes componentes del programa, con énfasis en la descripción de experiencias con el componente productivo en el estado de Oaxaca.

S/F

*Evaluación externa:* El Colegio de México propone una primera metodología general para la evaluación de Pronasol, coordinada por José Luis Reyna y Alberto García Rocha.

1994

*Administración pública:* la Secogef adquiere responsabilidades en materia de desarrollo administrativo y se convierte en Secretaría de Contraloría y Desarrollo Administrativo (Secodam).

1995

*Administración pública:* el gobierno de Ernesto Zedillo publicó el Programa para la Modernización de la Administración Pública Federal 1995-2000 (Promap), con la intención de desarrollar una nueva cultura del servicio público, guiada por criterios de eficacia, eficiencia y combate a la corrupción, y fundamentada en el NPM. Dicho programa identificaba entre los principales problemas a resolver: las deficiencias en los mecanismos de medición y evaluación del desempeño de la gestión pública,

---

*continúa...*

CUADRO 2  
(continuación)

---

para los que se proponía el diseño de un sistema basado en objetivos claros y metas cuantificables, indicadores de desempeño, la revisión y adaptación del Sistema Integral de Información por parte de la SHCP y la Secodam y el fortalecimiento y mayor independencia de los órganos internos de control preventivo y correctivo.

1995-2000

*Administración pública:* la SHCP fue encargada de estudiar la posibilidad de reformar el sistema presupuestal, enfocándolo a la evaluación de resultados de la gestión pública, a partir de la aplicación de un sistema de indicadores. Complementariamente, la Secodam se encargaría de los aspectos vinculados con el control de la calidad, la atención al cliente y la lucha contra la corrupción. De esta forma, al igual que en los casos de Uruguay y Chile, el presupuesto se volvía el instrumento de la reforma del Estado.

1995-2000

*Administración pública:* la SHCP construyó la denominada Reforma al Sistema Presupuestario, que proponía la integración de una Nueva Estructura Programática (NEP) y de un Sistema de Evaluación del Desempeño (SED) basado en resultados, con el fin de otorgar un amplio margen de libertad a los gerentes públicos. Pero, la propuesta anterior se instrumentó muy parcialmente, logrando sólo incorporar el discurso gerencialista sobre la reforma de la administración pública e introducir un poco más de claridad y transparencia en las cuestiones presupuestales (Arellano, 2003).

1995-2000

*Administración pública:* las aportaciones del Promap, en otros ámbitos, incluyeron la creación del Sistema Nacional de Quejas, Denuncias y Atención a la Ciudadanía, que promovió la participación de la última; y la reforma de un conjunto de leyes que permitieron modificar el funcionamiento de los Órganos de Control Interno de la Administración Pública Federal. Efectivamente, estos órganos dependían directamente de los titulares de cada institución, formando parte de las Oficialías Mayores o áreas similares, y se encargaban de realizar funciones de control y vigilancia de tipo correctivo, así

---

*continúa...*

CUADRO 2  
(continuación)

---

como evaluaciones enfocadas a la gestión de los procesos. A partir de las modificaciones introducidas, los mismos pasaron a ser designados y removidos por la Secodam, de la cual dependían funcional y jerárquicamente, logrando autonomía respecto de las entidades en las que desarrollaban su labor, y redireccionando sus evaluaciones hacia los resultados de la gestión.

1995-2000

*Evaluaciones institucionales:* hacia la segunda mitad de la década de 1990 se realizaron las primeras evaluaciones institucionales, de tipo interno, a programas de Liconsa y Diconsa.

1997

*Evaluaciones externas:* se inició el proyecto de evaluación externa más ambicioso solicitado por el gobierno mexicano, el del Programa de Educación, Salud y Alimentación (Progres, hoy Oportunidades).

1999

*Congreso de la Unión:* la CMH, luego de la modificación de los artículos 73, 74, 78 y 79 de la Constitución, fue sustituida en sus funciones por el órgano técnico, Auditoría Superior de la Federación (ASF), de la Cámara de Diputados del Congreso de la Unión, que continúa vigente a la fecha. Dicha modificación se realizó con la intención manifiesta de crear una instancia superior de revisión y auditoría, de carácter imparcial y que permitiera superar las limitaciones de su antecesora. Entre las últimas se planteaban las limitaciones impuestas a sus facultades para la revisión de Cuentas Públicas, la extemporaneidad de la emisión de resultados de sus revisiones, la ausencia de una garantía legal plena de su autonomía de gestión frente al Ejecutivo, la confusión generada por su propia denominación, el poco conocimiento por la ciudadanía de sus atribuciones y trabajos, etcétera.

2000

*Evaluaciones externas:* se difundieron los TR para la Evaluación Externa del Fondo de Coinversión Social 1999-2000, a cargo del entonces Instituto Nacional de Solidaridad (Insol).

---

*continúa...*

CUADRO 2  
(continuación)

---

DICIEMBRE DE 2000

*Congreso de la Unión*: se aprobó la Ley de Fiscalización Superior de la Federación que respalda jurídicamente a la ASF, la que cuenta con autonomía técnica, de gestión, material y financiera, para decidir respecto de su organización interna, funcionamiento y resoluciones. A diferencia de la CMH, su titular y funcionarios tienen la obligación de guardar secreto de su trabajo hasta tanto rindan los informes correspondientes [[www.asf.gob.mx](http://www.asf.gob.mx)].

2000

*Administración pública*: la Sedesol creó una Subsecretaría encargada de la evaluación de los programas que se encuentran bajo su responsabilidad.

2000

*Evaluaciones externas solicitadas por el Congreso de la Unión*: el Congreso decidió comenzar a exigir a partir de 2001 la evaluación externa anual de los programas financiados con recursos fiscales, sujetos a Reglas de Operación (RO).

2000

*Administración pública (local)*: se aprueba la Ley de Desarrollo Social del Distrito Federal (LDSDF).

INICIO DE 2001

*Administración pública*: se instaló la Comisión Intersecretarial para la Transparencia y el Combate a la Corrupción, que además de su obvio objetivo, busca elevar los niveles de eficiencia, eficacia y calidad en el servicio público. En abril de ese mismo año se aprobó el Programa Nacional para la Transparencia y el Combate a la Corrupción, que entre sus líneas estratégicas incluye: mecanismos de rendición de cuentas, de supervisión y control, y de participación ciudadana.

JUNIO DE 2001

*Administración pública*: se implementó el Programa de Corto Plazo para el Fortalecimiento del Control Interno dentro de las dependencias y entidades de la Administración Pública Federal, con el objetivo

---

*continúa...*

CUADRO 2  
(continuación)

---

de promover la autonomía de gestión de las instituciones y reducir la discrecionalidad en el manejo de los recursos, así como permitir una mayor transparencia en el ejercicio del presupuesto.

2002

*Información para la evaluación por la sociedad civil:* en junio se aprueba la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental y la Comisión Federal de Mejora Regulatoria (Cofemer) se convierte en el primer órgano público federal que rinde cuentas a la ciudadanía sobre su gestión.

2002

*Administración pública:* se expide el “Acuerdo por el que se expiden los requisitos mínimos que las dependencias y entidades de la Administración Pública Federal deben considerar para seleccionar a las instituciones académicas y de investigación u organismos especializados, de carácter nacional o internacional, interesados en realizar la evaluación de resultados de los programas gubernamentales sujetos a RO en el ejercicio fiscal 2002”, firmado por la SHCP y la Secodam.

2002

*Administración pública:* se establece el Sistema de Seguimiento de Metas Presidenciales (SIMEP), bajo la responsabilidad de la Secodam y para uso interno del Gobierno Federal.

2003

*Administración pública:* la Secodam se transforma en Secretaría de la Función Pública (SFP), continuando con su responsabilidad de fortalecer los sistemas de gestión y control, combatir la corrupción y fomentar la transparencia de la Administración Pública Federal, entre otros [[www.sfp.gob.mx](http://www.sfp.gob.mx)].

2003

*Administración pública:* se aprueba la Ley General de Desarrollo Social (LGDS) que prevé la creación de un Consejo Nacional de Evaluación integrado por ocho personas, seis de las cuales deben ser investigadores académicos que cumplan determinados requisitos que garanticen su calidad, y se

---

*continúa...*

CUADRO 2  
(continuación)

---

reconoce a la Contraloría Social como el mecanismo en que los beneficiarios verifican la aplicación de recursos y el cumplimiento de metas.

2004

*Organizaciones de la sociedad civil*: se publica la Ley Federal de Fomento a las Actividades Realizadas por Organizaciones de la Sociedad Civil, que en su artículo 6, inciso XI, promueve su mayor participación en los procesos de políticas públicas, en particular en relación con el seguimiento de las acciones públicas realizadas.

2004

*Administración pública*: se diseñan indicadores de desempeño bajo el liderazgo de la SFP mediante el Modelo Integral de Desempeño de Órganos de Vigilancia y control (MIDO).

2006

*Administración pública*: se expide el Reglamento de la LGDS y se crea el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval).

2006

*Administración pública* (local): se reglamenta la Ley de Desarrollo Social del Distrito Federal.

2006

*Administración pública*: la Ley de Presupuesto y Responsabilidad Hacendaria formaliza el Presupuesto por Resultados.

2007

*Administración pública*: la SHCP presenta a la Cámara de Diputados su propuesta de Sistema de Evaluación del Desempeño (SED) y el Coneval expide los “Lineamientos generales para la evaluación de programas federales de la Administración Pública Federal”, que sustituirán al Acuerdo de SHCP-Secodam de 2002.

---

*continúa...*

CUADRO 2  
(termina)

---

2007

*Administración pública* (local): se decide la creación del Consejo de Evaluación del Desarrollo Social del DF (Evalúa DF) como organismo público descentralizado, con personalidad jurídica y patrimonio propio, así como autonomía técnica, de gestión y presupuestaria.

2008

*Administración pública*: se expide la Ley de Transparencia y Acceso a la Información Pública del DF y se integra Evalúa DF.

2008

*Administración pública* (local): se integra Evalúa DF con seis ciudadanos académicos, cuatro de los cuales deben pertenecer al Sistema Nacional de Investigadores para garantizar su calidad.

---

Fuente: elaboración propia.

De esta forma, durante la mayor parte del siglo pasado, a nivel gubernamental, más que evaluaciones del impacto de las políticas y programas, encontramos mecanismos de control jurídico y financiero del gasto (actividades de contraloría) y seguimientos de avance de metas. Cuando existían recortes presupuestales, las primeras actividades que se reducían eran las de evaluación, consideradas menos útiles que las sustantivas.

En resumen, puede decirse que, hasta 1999, las actividades de control y evaluación de la administración pública eran ejercidas por la SHCP, la entonces Secodam y la CMH de la Cámara de Diputados. Se trataba fundamentalmente del control del gasto público, la legalidad de los procesos, el combate a la corrupción, etcétera, pero no se tiene conocimiento de que estas instancias evaluaran algún programa gubernamental, menos social, en términos de sus resultados o impactos.

La SHCP era, y continúa siendo, el órgano financiero del Estado Federal encargado de los asuntos relativos a los ingresos de la Federación y entidades paraestatales; del proyecto y cálculo de egresos del Gobierno Federal (GF) y

de la administración pública paraestatal; de la planeación, presupuestación, contabilidad y evaluación; entre otras.

La Secodam era la encargada de planear, organizar y coordinar el sistema de control, y de realizar la evaluación gubernamental, la inspección del gasto público federal en congruencia con su presupuesto de egresos, el establecimiento de bases para regular las auditorías a dependencias y entidades de la administración pública y efectuarlas, el control de que el enriquecimiento de los funcionarios estuviera debidamente justificado, que las obras del gobierno avanzaran al ritmo acordado y tuvieran la calidad convenida, etcétera. Como ya se mencionó, en el sexenio de Ernesto Zedillo se le agregó la responsabilidad del desarrollo administrativo de la administración pública.

La CMH ejercía el control legislativo sobre el manejo que la Administración Pública realizaba de los ingresos y egresos que autorizan las leyes y presupuestos y, en general, sobre la actividad financiera y patrimonial.

En la actualidad, las actividades internas de control y evaluación de la administración pública, están a cargo de: la SHCP, que continúa con su responsabilidad en materia financiera; la SFP, que persigue el fortalecimiento de los sistemas de gestión y control, el combate a la corrupción y el fomento de la transparencia de la administración pública federal; y la ASF de la Cámara de Diputados, que fiscaliza la cuenta pública. Sin embargo, la decisión del Congreso para que se realicen evaluaciones externas a los programas que reciben recursos fiscales, iniciada en el año 2001,<sup>4</sup> provocó un enorme aumento en la cantidad de estos trabajos, realizados mediante la contratación de consultorías privadas de evaluación externa, que enfrentaron nuevas restricciones técnicas y financieras. El BM reconoce los esfuerzos realizados por el gobierno para mejorar especialmente el monitoreo y la evaluación de los programas sociales (BM, 2004a).

Compartimos la idea de que:

La evaluación gubernamental sólo es posible en un sistema político donde el debate parlamentario está enriquecido por posiciones partidistas distintas

<sup>4</sup> El BM (2004b) señala que la decisión se tomó en 1999 para aplicarse al siguiente, pero en nuestra revisión del PEF 2000 no encontramos esa referencia, que en cambio aparece en el PEF 2001.

del partido que tiene el poder, pero sobre todo, por una opinión pública concientizada, cuyo papel en los intercambios gobierno-sociedad sea tomado seriamente en cuenta (Pardo, 2000:250).

Contamos con la primera condición, pero aún falta bastante para la segunda. Por otra parte, la evaluación realizada por el Poder Legislativo sólo resulta garantía de objetividad si se crean o fortalecen los apoyos técnicos que le permitan realmente analizar la información que reciben.

Obviamente, falta desarrollar un mecanismo efectivo de participación ciudadana en estos procesos, que demuestre que la evaluación no sólo responde a la escasez de recursos y racionalización del gasto, sino que el gobierno se preocupa también por impulsar la transparencia y el desarrollo de la democracia. Incluso las organizaciones sociales y civiles mexicanas han estado impulsadas por su voluntad de contribuir a la resolución de problemas y han considerado durante mucho tiempo a la evaluación como una forma de distraer sus recursos financieros y humanos en actividades no sustantivas. A esto se agrega el rechazo que tradicionalmente generan las evaluaciones provenientes de las fundaciones y organismos que aportan su financiamiento, lo que provocó que se las concibiera más como un control externo no deseado, que como un proceso propio que puede ayudar al aprendizaje de los participantes y la comunidad en su conjunto. Dichas razones han contribuido para que hasta fechas muy recientes se preocuparan poco de capacitarse en esta materia y, en todo caso, prefirieran las evaluaciones cualitativas, que si bien permiten profundizar en casos concretos no nos ofrecen una visión agregada de conjunto.

### **Profesionalización de la evaluación**

Cunill y Ospina mencionan el necesario desarrollo de capacidades en América Latina, a partir de un proceso interactivo, gradual y creativo de aprendizaje en que tanto los sujetos individuales como colectivos mejoran y/o complementan sus conocimientos, destrezas, actitudes y sentimientos. Este último objetivo (modificar sentimientos) es poco común que se mencione explícitamente a pesar de su enorme trascendencia vinculada con el compromiso público, la voluntad y la energía con que se desarrolla la gestión; la razón es que existen serias

dificultades pedagógicas en torno a cómo lograrlo, muy ligadas a la enseñanza de valores y demás aspectos éticos.<sup>5</sup>

Cardozo (2006a) sólo identificó la existencia de un programa de maestría en Evaluación en Costa Rica. En la actualidad, la situación es similar; sin embargo, identificamos un curso virtual que se ofrece en Argentina<sup>6</sup> y una pequeña maestría virtual que la ReLAC estaba diseñando para América Latina en 2007. La misma se impartió exitosamente en 2008, como prueba piloto, en la Facultad Latinoamericana de Ciencias Sociales (Flacso) de Chile, pero se desconocen los motivos por los cuales el proyecto parece haberse detenido.

A nivel internacional, y accesibles en América Latina, han existido algunos cursos virtuales intermitentes como “Marco lógico para el diseño de proyectos”, “Monitoreo y evaluación de proyectos” y “Evaluación de impacto ambiental” del BID; e “Indicadores y metodologías cualitativas para la gerencia social” y “Monitoreo de los programas sociales” del Instituto de Desarrollo Social del BID (Cardozo, 2009).

También en México se percibe poca preocupación universitaria y de otras instituciones educativas y de investigación por el tema. Si bien fueron éstas las que iniciaron la realización de evaluaciones independientes, a la fecha no existen programas especialmente enfocados al desarrollo de esta incipiente profesión<sup>7</sup> y puede constatarse el reducido número de tesis vinculadas con el tema, que en los últimos años se han venido incrementado. Además, existe una enorme carencia de material didáctico propio: la inmensa mayoría de los libros y revistas que exponen métodos de evaluación y presentan experiencias de aplicación, provienen especialmente de Estados Unidos y, en mucho menor medida, de Europa o Canadá; respondiendo a contextos muy diferentes de los nuestros. Pocas fueron las evaluaciones hechas desde la academia antes del año 2000 y, a nivel institucional, el seguimiento fue mínimo, no difundido y a menudo incluso se limitó su acceso al investigador.

<sup>5</sup> Consultar curso del BID “Cómo enseñar ética, capital social y desarrollo en la universidad? En [www.iadb.org/etica](http://www.iadb.org/etica).

<sup>6</sup> El curso se denomina “Evaluación de resultados e impactos de organizaciones y programas públicos” y es impartido por el Centro de Desarrollo y Asistencia Técnica en Tecnología para la Organización Pública AC. (TOP) y la Universidad del Litoral.

<sup>7</sup> El ITAM es el único que cuenta con un centro especializado en materia de evaluación, pero se trata de un enfoque *ex-ante*, especialmente dirigido a proyectos de desarrollo regional.

Los esfuerzos más recientes en este campo incluyen, entre otros: un primer diplomado en evaluación organizado por el Centro de Estudios para el Desarrollo Sustentable y la Soberanía Alimentaria (CEDRSSA) de la Cámara de Diputados y otro promovido por la Universidad Autónoma Metropolitana, Unidad Xochimilco (UAM-Xochimilco), ambos en 2006; otro diplomado y diversos talleres de capacitación organizados por Evalúa DF en 2008-2009, y un nuevo diplomado impartido en 2009-2010 por la Escuela de Administración Pública del DF. Por otra parte, Coneval junto con CEPAL y, por otro lado, la SHCP, han capacitado intensamente en torno al diseño del Marco Lógico.

### **Comentarios finales**

El capítulo nos permite comprender las dificultades técnicas y políticas encontradas en América Latina para impulsar procesos de evaluación, que como se vio en el capítulo previo, no son exclusivas de esta región ni producto, al menos único, del nivel de desarrollo, así como la contribución de los organismos internacionales al desarrollo de la misma.

Colombia y Costa Rica se adelantan, a inicios de la década de 1990, en el establecimiento de un modelo de evaluación de programas, con la intención de reforzar las decisiones de políticas ligadas al Poder Ejecutivo; mientras Chile y Uruguay buscan resultados medibles para orientar las decisiones presupuestales de sus Poderes Legislativos.

El caso de México aparece como particularmente lejano de Estados Unidos y Reino Unido debido a la importancia que en ellos adquiere el Poder Legislativo en la contratación de evaluaciones tendientes a vigilar el desempeño del Ejecutivo y a la amplia disponibilidad de recursos financieros y profesionales para su realización. De Estados Unidos lo separan además, unas cinco décadas en la institucionalización del sistema de evaluación de programas.

En cambio, México se muestra más cercano a la situación descrita en Francia antes del año 2000, caracterizada por la tradición de centralismo y la resistencia a la evaluación, el Poder Legislativo al margen del proceso, retraso en los procesos de rendición de cuentas, falta de desarrollos metodológicos propios, medición de resultados más que de impactos, y limitada utilización de los resultados. No obstante, el caso francés se distanciaba de la evaluación basada en la experimentación para

buscar un modelo propio, que respondiera al incremento de la democratización y el pluralismo de actores, lo que no ha sucedido en el caso mexicano. Como hemos visto, actualmente Francia muestra una tendencia a regresar a los modelos experimentales, pero también a los métodos multicriterios.

En el contexto latinoamericano, México avanza en la generalización de la evaluación de sus programas sociales (el resto lo hace en forma selectiva) y en el esfuerzo aplicado en el caso particular del programa de Desarrollo Humano Oportunidades. Comparte con los demás países de la región los problemas de sus sistemas de información, medición de impactos, capacitación, aprendizaje social y utilización de los resultados, así como el apoyo recibido de organismos internacionales como el BM o el BID. En cambio, no cuenta aún con una cultura de la evaluación orientada a resultados como la mencionada en Colombia, una difusión de los informes que mantenga informada a la ciudadanía como estaría sucediendo en el caso de Costa Rica y procesos de capacitación apoyados con la preparación de guías de evaluación como se hizo en Reino Unido, Francia y los cuatro países de América Latina mencionados.

La observación realizada en México muestra que la evaluación actual de políticas y programas públicos federales cuenta con poca información publicada hasta el presente sobre los procesos de evaluación emprendidos, en particular: 1) los métodos y técnicas aplicados, 2) la calidad de los resultados alcanzados, y 3) el nivel de utilización de dichos resultados en los procesos decisorios. Se trata entonces de responder a preguntas como: ¿por qué, qué y cómo se evalúa?, ¿quién lo hace?, ¿cuánto cuesta?, ¿qué problemas enfrenta?, ¿para qué sirve? Si bien este panorama se ha venido mejorando, aún es insuficiente para promover la participación ciudadana en los procesos de políticas públicas, contribuir a un proceso de aprendizaje, lograr más transparencia en las decisiones sobre el uso de recursos, eficacia en su aplicación a los objetivos perseguidos y pluralidad para contribuir a la democratización del país.

También es importante resaltar en México la aparición de Consejos locales de evaluación en el Gobierno del Distrito Federal (GDF) y, más recientemente, en los estados de México y Puebla, que no todos siguen la orientación encabezada por el Coneval a nivel federal.

Finalmente, resaltan las pocas oportunidades de profesionalización en materia de evaluación en los países revisados y especialmente en México.

Para responder a las preguntas previamente planteadas, en el siguiente capítulo se analizarán los aportes que puede realizar la aplicación de metaevaluaciones a las evaluaciones realizadas.

## CAPÍTULO VI

### Evaluación de programas sociales en México

En el presente capítulo corresponde exponer las particularidades que asume la evaluación aplicada a este tipo particular de la acción pública (social) y revisar los primeros intentos por evaluar algunas de ellas.

Como se indicó en el capítulo anterior, muchos fueron los programas sociales aplicados por México durante el siglo XX; entre los más recientes se pueden mencionar el Programa de inversiones públicas para el desarrollo rural (Pider) durante el gobierno de Luis Echeverría (1970-1976), la Coordinación del Plan general de zonas deprimidas y grupos marginados (Coplamar) y el Sistema alimentario mexicano (SAM) en el sexenio encabezado por José López Portillo (1976-1982), el Programa de nutrición y salud de Miguel de la Madrid (1982-1988), el Programa Nacional de Solidaridad (Pronasol), creado por Carlos Salinas de Gortari (1988-1994) y el Programa de Educación, Salud y Alimentación (Progresá), emprendido por Ernesto Zedillo, antecedente inmediato del actual programa Oportunidades.

Salvo los dos últimos, los demás tienen en común que no se les realizó ninguna evaluación relevante. En cambio, en el periodo 1995-2000, se tiene conocimiento de algunos trabajos efectuados en el marco de los programas a cargo de Liconsa, Diconsa o Insol, si bien no existen documentos públicos que permitan analizar esas experiencias. A partir de 1998, en los casos de Pronasol y Progresá se cuenta con diversos trabajos vinculados con su evaluación, que resultan vitales para entender el tipo de evaluación que se está realizando actualmente en Oportunidades, y en general, en los programas de desarrollo social.

El capítulo termina con la exposición de los procedimientos instalados a partir de 2001.

## Particularidades de la evaluación de políticas sociales

A pesar de la importancia de los mecanismos de descentralización, externalización y focalización implementados en América Latina, hace ya 10 años, Cohen (2001) manifestaba que no existía claridad sobre los beneficios de su aplicación.<sup>1</sup> Su hipótesis era que esto se debía al traslado acrítico de estrategias de intervención que habían sido exitosas en países europeos industrializados que intentaban avanzar en formas de organización y gestión “posburocráticas”, cuando América Latina aún estaba permeada por todos los vicios “preburocráticos” y no existían condiciones que permitieran recrear y regular los escenarios de mercado para el buen funcionamiento de la externalización. El BM (2004) también reconoció la falta de información sobre el impacto causado por los programas sociales en los países en desarrollo.

Cohen (2001) resalta la importancia de proceder a su evaluación considerando: 1) la magnitud del gasto social, 2) el porcentaje del gasto social efectivamente redistributivo, 3) la eficiencia en la aplicación del gasto anterior, y 4) su impacto en la solución del problema.

Las razones de esta secuencia son obvias: no puede haber resultados relevantes en programas que carecen de recursos suficientes; no basta con gastar, pues muchos recursos pueden insumirse en los procesos sin llegar a la población objetivo;<sup>2</sup> y, suponiendo que lleguen, por lo general se desconoce su efectividad para resolver o aliviar el problema que ocasionó el programa, el grado en que justifican su costo y quienes son sus verdaderos beneficiarios o perjudicados.

<sup>1</sup> Con base en un trabajo realizado en la Comisión Económica para América Latina (CEPAL), en 1998, plantea que los beneficios reales en términos de impacto y eficiencia derivados de la introducción de las dos primeras estrategias citadas en cinco programas sociales de la región latinoamericana, serían marginales o incluso nulos, cuando no se implementan correctamente.

<sup>2</sup> Se refiere un trabajo que demostraría en el caso chileno que sólo 13% del gasto social total llega al 30% más pobre de la población; situación que se considera se repite en otros contextos nacionales de América Latina. Otro trabajo realizado en seis países latinoamericanos reporta que 90% del gasto público social corresponde a gasto corriente. En el caso del Progreso mexicano, su evaluación calculó que los gastos administrativos o programáticos ascendían a 8.2%, los relacionados a la focalización del programa a 30% y los vinculados a su condicionalidad, 26%; lo que fueron considerados porcentajes pequeños en comparación con programas mexicanos previos y otros dirigidos a grupos particulares en países extranjeros.

Considerando que los programas sociales se caracterizan por: requerir del mediano y el largo plazo para alcanzar sus objetivos redistributivos, por ser afectados por múltiples variables contextuales, una gran variedad de actores participantes con intereses en conflicto y una fuerte atribución de importancia a la participación social, a la transparencia y la responsabilidad de la gestión, al menos en el discurso (Cardozo, 2000), su especificidad debe ser tomada en cuenta en su evaluación.

Nadie parece negar que la efectividad de un programa social depende también del grado en que éste recupera las necesidades sociales reales y responde a dichas circunstancias, lo que implica disponer de un diagnóstico que priorice los problemas a atender (Rossi y Freeman, 1993), que se apliquen mecanismos de control y evaluación desde el inicio, y que la información generada sea puesta a disposición del público para aumentar la transparencia y garantizar su participación (Franco, 2003). En el informe sobre la lucha contra la pobreza en México, el BM (2004) identificó como uno de los tres retos pendientes: fortalecer el monitoreo y la evaluación de impacto de los programas sociales.

En el caso de la evaluación de la participación social, estrategia ampliamente considerada relevante en el marco de los programas sociales, Bichmann y otros (1989) señalan que ésta puede ser estudiada tomando en cuenta diferentes áreas donde la misma puede incidir: orientación, liderazgo, organización, recursos y gestión. Cada uno de estos capítulos se califican y sus resultados se grafican con base en las respuestas obtenidas a un conjunto de preguntas aplicadas a los diferentes actores sociales.

En este trabajo se considera imprescindible la complementación de los agregados económicos planteados con indicadores cualitativos de pertinencia, coherencia, satisfacción, calidad, participación social, etcétera.

Además, es importante resaltar que la evaluación de políticas públicas es mucho más compleja que la de programas públicos, especialmente en el área social, ya que “sus metas son variadas, sus actores numerosos, los instrumentos diversificados, los resultados a menudo decepcionantes” (Kessler y otros, 1998, traducción libre). Según Perret (1993), algunos países se refieren a la evaluación en ambos campos (sin hacer diferencias entre éstos en Reino Unido y haciéndolas en Estados Unidos); otros sólo mencionan al campo de programas públicos (Canadá, Australia, Alemania); otros más, sólo a las políticas como conjunto de programas (Países Bajos), o bien, a la evaluación de políticas cuando en realidad se

hace evaluación de programas (Francia). La situación mexicana se acerca bastante a esta última; sin embargo, retomando los dos conceptos planteados en el capítulo II, Evalúa DF ha tratado de introducir la diferencia evaluando programas en 2008 y políticas en los años subsiguientes, para lo cual produjo una “Nota de orientación para la evaluación de las políticas de desarrollo social del GDF” [[www.evalua.df.gob.mx/convocatoria/2010/nota\\_met\\_may\\_10\\_def.pdf](http://www.evalua.df.gob.mx/convocatoria/2010/nota_met_may_10_def.pdf)].

### **Evaluaciones de Pronasol**

Pronasol inició actividades en 1988, siendo el primer programa mexicano dirigido a combatir la pobreza en forma focalizada, privilegiando a las comunidades organizadas.

Entre 1990 y 1994, el programa fue objeto de múltiples intentos de evaluación parcial, ya sea a nivel estatal y/o enfocados a aspectos específicos del mismo. Pero, pocos fueron los casos en que se intentó su evaluación integral a nivel nacional. A continuación se exponen tres antecedentes relevantes en este último sentido:

*a)* Como ya se mencionó en el capítulo anterior, Székely (1993) expresó la ausencia de elementos en la literatura que le permitieran responder a las preguntas formuladas por la OIT para proceder a su evaluación, salvo en lo relativo a los recursos ejercidos por el programa, su relación con el PIB y el gasto público. Las preguntas no respondidas se referían a: la identificación de efectos compensatorios y redistributivos, el nivel de focalización hacia los pobres, los efectos productivos en materia de creación de empleo y empresas, en la construcción de infraestructura y dotación de servicios, y a la evaluación de la organización y operación institucionales. Concretamente, expresó que “existen datos respecto a cantidades invertidas y/o volúmenes de obra, pero no evaluaciones propiamente dichas, en términos de costo beneficio, ni frente a metas numéricamente preestipuladas de reducción en los déficits que marcan los índices de pobreza” (Székely, 1993:32). Por esa razón, insistió en que su trabajo se basaba en su apreciación sensorial de la situación, enriquecida con el relato de algunas experiencias.

Székely procedió entonces a partir del análisis y relación entre los tres componentes del programa: asistencial, participativo y competitivo. En el primero resaltó el esfuerzo financiero del programa y lo comparó con los recursos

que serían necesarios para superar las carencias de los sectores más pobres, tomando a la política de reactivación económica general como un dato.

En relación con el componente participativo argumentó la fundamentación de la focalización hacia los pobres organizados, en lugar de hacerlo hacia los más pobres, con base en un criterio de tipo costo/beneficio. Además insistió en la ventaja de abandonar los canales burocráticos y corporativizados de circulación del gasto social.

En el tercero recurrió a la narración de experiencias de programas productivos en el estado de Oaxaca: Apoyo a cafeticultores, Fondos regionales de solidaridad para el desarrollo de los pueblos indígenas, Fondo para el desarrollo pesquero del litoral oaxaqueño y Fondo nacional de empresas de solidaridad.

En todos los casos retomó los datos numéricos disponibles en tanto aproximaciones para apoyar sus reflexiones cualitativas. También puso énfasis en analizar las críticas que se le hacían al programa.

Concluyó en la urgencia de ampliar el programa mediante un esfuerzo económico y político sin precedentes, acompañado de una modificación de las funciones ejercidas por las instituciones y las bases sociales organizadas.

El intento de evaluar el Pronasol en su conjunto dejó clara la insuficiente información disponible para la realización de un trabajo riguroso que pretendiera basarse en el seguimiento de metas y medición de impacto.

*b)* Un segundo intento fue realizado por Jaso y Merino (1994), quienes procedieron a realizar una síntesis evaluativa y crítica de los trabajos parciales realizados, cuyas conclusiones resultaban a menudo complementarias y en otros casos definitivamente contradictorias. Consideraron al programa como una política pública con propósitos definidos, que ameritaba una evaluación lo más objetiva posible, al margen de los intereses políticos del gobierno y los partidos. Esto implicaba dificultades importantes debido a: el limitado avance de la evaluación de políticas públicas, la complejidad y novedad de la organización del programa, la importancia atribuida a la organización social, la diversidad de métodos empleados en las evaluaciones parciales analizadas, las diferentes fechas en que fueron realizadas, etcétera. Resolvieron entonces ofrecer una visión, de énfasis cualitativo, que recogiera todas las opiniones con su respectivos fundamentos y evidencias, tratando de rescatar sus ideas centrales.

Dichas ideas indujeron la estructura final de su trabajo: dos grandes capítulos sobre la operación del programa y la organización social y otros dos complementarios sobre su impacto económico y político.

El primero revisa los objetivos del programa, su organización, mecanismos de implementación y resultados verificables, señalando aciertos y deficiencias. Concluye que las evaluaciones coinciden en destacar la complejidad ejecutiva del Pronasol y su operación centralista. Agregan que los recursos no siempre fueron destinados con precisión a los grupos más pobres, ni a los proyectos que podrían ayudarlos más.

El capítulo sobre organización social, considerada la diferencia específica de Pronasol, analiza la integración y funcionamiento de los Comités de Solidaridad que aportaron dinero y mano de obra al programa, el perfil socioeconómico de los participantes, su nivel de satisfacción y expectativas de continuidad. Se concluye que la misma impulsó una relación diferente entre el gobierno y la sociedad, potenciando formas tradicionales de cooperación comunitaria, entre las que sobresalen los resultados de participación obtenidos por los comités de Escuela digna.

En materia de impactos económicos el trabajo discute la confiabilidad de los datos INEGI-ONU-CEPAL que referían una disminución de la pobreza externa de 14.9 a 13.6 millones de personas entre 1989 y 1992. Concluye que el programa fue eficaz en el mejoramiento de la calidad de vida de los grupos más pobres cuando ésta se mide en términos de acceso a satisfactores colectivos, pero ha sido insuficiente para garantizar los ingresos familiares necesarios para la satisfacción de sus necesidades esenciales.

La crítica más frecuente al programa es que habría sido utilizado para favorecer electoralmente al Partido Revolucionario Institucional (PRI), pero los aspectos políticos del mismo fueron los más debatidos y con menor información empírica disponible. Sin embargo, el trabajo afirma que los datos existentes sobre distribución del gasto no muestran evidencias de que Solidaridad haya reorientado regionalmente la misma en función de las elecciones locales.

Se concluye resaltando los aciertos del programa pero también su insuficiencia para revertir las inercias de funcionamiento y las causas estructurales de la pobreza.

c) En tercer lugar, con la intención de contribuir al desarrollo de una metodología específica para la evaluación del impacto socioeconómico del programa, el Colegio de México realizó un trabajo coordinado por Reyna y Rocha (s/f), en el que también participó Székely como uno de los investigadores del equipo. El mismo constituye un antecedente obligado de lo que serían las posteriores evaluaciones de programas sociales.

Dicho trabajo define ante todo un marco de referencia para la evaluación, comenzando por identificar las hipótesis subyacentes en el programa: la población objetivo llegará a alcanzar mayores ingresos al mejorar su estado de salud, adquirir mayores conocimientos, contar con mayor infraestructura de servicios y aumentar o mejorar su capacidad de participación en la toma de decisiones. Sin embargo, esta formulación implica aún otra hipótesis que el documento no explicita: que la obtención de mayores ingresos redundará en mejores condiciones de vida para la población (objetivo final, señalado en primer lugar, por el programa).

El documento sugiere tres aspectos interdependientes a ser considerados en la evaluación: la adecuación del programa a las necesidades y demandas de la población, la organización interna del programa y su relación con su población objetivo, y el impacto en la generación de un ingreso adicional para la última.

Estos aspectos y la congruencia entre el objetivo final y las estrategias del programa, indicada en el punto inicial, forman parte de los temas de diseño, funcionamiento y medición de impactos que este trabajo considera imprescindibles, pero poco considerados en los TR e informes de evaluación de 2002, que resaltan más los recursos aplicados, las metas alcanzadas, la cobertura, los resultados logrados, y el apego de la operación a la normatividad.

Reyna y Rocha reconocen la heterogeneidad de estrategias del programa, los distintos tiempos de avance de los proyectos y las dificultades para identificar tanto los impactos esperados como los imprevistos, los inmediatos como los mediatos, por lo que sólo proponen una metodología de “aproximación” a los impactos del Pronasol, por estados y basada en dos dimensiones de análisis: los objetivos propios de cada estrategia y las características organizativas y de participación de la población.

En los programas sociales subsiguientes se abandonará este énfasis de Pronasol en la organización de la población; sin embargo, el discurso de las políticas seguirá insistiendo en la participación de la misma, aunque descuidando los canales para su implementación y evaluación.

La propuesta de metodología de evaluación analizada sugiere considerar diversos conjuntos de indicadores: de medición operativa de los niveles de pobreza extrema, comunes a todos los componentes, y específicos para cada uno según sus campos de acción (salud, educación, infraestructura, etcétera).

Actualmente, casi todos los programas se limitan a la medición específica de sus objetivos y metas inmediatos. La medición de la evolución de la pobreza se hace por otros medios (encuestas del INEGI, estudios que utilizan su información, Comité técnico para la medición de la pobreza desde 2001) y resulta prácticamente imposible saber si todos los programas aplicados han contribuido a la misma y en qué proporción, por lo que suele correlacionarse sólo con el mayor de los programas (Oportunidades).<sup>3</sup>

El documento del Colmex plantea el conjunto de preguntas a responder, ofrece una propuesta de indicadores y señala las relaciones entre los mismos, sugiere la forma de evaluar la participación de cada actor relevante del programa, la manera de determinar la muestra estadística para la aplicación de encuestas, y la forma de presentar los resultados.

El capítulo relativo a la medición del cambio en los niveles de pobreza es prácticamente indiscutible desde el punto de vista teórico: se trata de calcular un índice antes y después de las acciones de Progresá y medir la relación costo-beneficio como el cociente entre el costo total de las acciones y el cambio logrado en el nivel de bienestar. Sin embargo, las seis páginas de extensión del capítulo no abordan la complejidad que implica la identificación y el cálculo del índice de pobreza y tampoco plantean el riesgo de adjudicar al programa efectos provocados por otras variables y los métodos disponibles para su reducción (experimentación con grupo de control).

Otro breve capítulo de gran interés lo constituye el estudio de la relación entre los indicadores de impacto y la participación. Actualmente se cuenta con pocos casos de participación significativa de los beneficiarios pero, para no promover o rechazar la misma sin evidencia empírica, debería diseñarse un experimento controlado para determinar si la misma tiene o no efectos positivos y en qué condiciones.

<sup>3</sup> El documento del BM (2004b), en cambio, plantea que no se sabe en qué porcentaje la pobreza ha disminuido por los programas gubernamentales aplicados y por las remesas recibidas por los mexicanos desde el exterior.

La presentación de resultados se limita, en dos cuartillas, a indicar que se dará respuesta a las preguntas planteadas. Como puede verse, el documento no profundiza en las grandes dificultades que en la operativización de la metodología de evaluación se plantean, pero constituye un primer esbozo de cómo intentarla que se acompaña de valiosos anexos (indicadores, cuestionarios, ejemplos de medición de pobreza y de cambios en el nivel de bienestar, etcétera) y señala aspectos problemáticos que aún hoy no han sido plenamente resueltos.

### **Evaluaciones de Progresá 1998-2000**

Progresá constituye el antecedente inmediato del actual Programa de Desarrollo Humano Oportunidades. Estuvo bajo la responsabilidad de la Sedesol, y contó desde 1997 con una Coordinación Nacional (Conprogresá) integrada por los titulares de la secretaría anterior, quien lo presidía, y de la Secretaría de Salud (Ssa), la Secretaría de Educación Pública (SEP) y la Secretaría de Hacienda y Crédito Público (SHCP).

Dicho programa incluyó tanto transferencias de efectivo a las madres de familia para mejorar los niveles de consumo de los pobres en el corto plazo, como inversiones en capital humano para lograr su desarrollo social, mismas que dieron inicio en mayo de 1998. Mientras Pronasol, su antecesor, había puesto el énfasis en la organización de las comunidades, Progresá se diseñó para dirigirse a individuos y, después de su discusión, se resolvió reorientarlo hacia las familias. De esta forma el programa procedió con una doble focalización: a las localidades más marginadas y, dentro de éstas, a los hogares más pobres. Este diseño implicó que los hogares muy pobres ubicados en localidades no marginadas quedaran fuera de las metas del programa.

Fue el primer programa mexicano (y tal vez el único) cuya evaluación de impacto fue prevista desde su diseño,<sup>4</sup> partiendo de un diagnóstico de la situación antes de su inicio, medida a partir de la aplicación de encuestas a hogares de octubre-noviembre de 1997 y marzo de 1998; y aplicando tres evaluaciones posteriores en octubre-noviembre de 1998, junio y noviembre-diciembre de

<sup>4</sup> En esto habría influido en parte la presión del BM.

1999. La misma revisó los impactos en materia de educación, nutrición, salud y pobreza rural, así como la operación general del programa. Posteriormente se le hicieron ajustes y se continuó con sus evaluaciones, las que en 2003 habían generado once volúmenes de documentos.

Si bien en la evaluación inicial participaron organizaciones mexicanas como la Universidad Iberoamericana (UIA), el Instituto Nacional de Salud Pública (INSP) y la Escuela Nacional de Antropología e Historia (ENAH), la responsabilidad del proceso estuvo a cargo del International Food Policy Research Institute (IFPRI) de Washington, DC., mediante contratación directa y bajo la dirección del doctor Emmanuel Skoufias, quien posteriormente pasó a formar parte del BID.<sup>5</sup> O sea que el gobierno no le confió la responsabilidad de esta primera evaluación a los propios connacionales, sino que recurrió al empleo de expertos de un instituto internacional de investigación.

El gobierno de Ernesto Zedillo pagó por esta evaluación \$4.35 millones en 1998, \$11.31 en 1999 y \$12.30 en 2000. Este monto no cubría el trabajo de campo, del que se encargaba el mismo programa.

La evaluación del Progreso utilizó un diseño experimental con la finalidad de garantizar un alto rigor en la identificación de relaciones causales entre el programa y los impactos identificados, con mediciones “antes-después”, talleres en los siete estados piloto y discusión en grupos focalizados (médicos, enfermeras, promotoras, profesores, etcétera), además del empleo de técnicas tradicionales como encuestas y entrevistas.

Dicho diseño hubiera requerido idealmente una selección aleatoria de beneficiarios y no beneficiarios (grupo de control), a nivel individual, pero debido a que la rigurosidad es costosa, se optó por una selección aleatoria a nivel de localidades, que no garantiza de igual manera la ausencia de sesgos en la medición de impactos.<sup>6</sup> Las localidades seleccionadas fueron 506, compuestas

<sup>5</sup> Véase al respecto Skoufias (2001 y 2003) [[www.ifpri.org](http://www.ifpri.org)]. También los ocho tomos publicados en el año 2000 con las evaluaciones realizadas en 1998-2000.

<sup>6</sup> Consultar en [www.ifpri.org](http://www.ifpri.org), Behrman y Todd (febrero de 1999), en su reporte a Progreso “Randomness in the experimental samples of Progreso”, en el que manifiestan que no hay diferencias significativas en las características generales de edad, educación, ingreso o acceso a la salud entre el grupo de tratamiento y el de control; sin embargo, a nivel individual y familiar,

por 302 de un total de 4 546 en que el programa fue aplicado en su fase II y 186 de 1 850 en que esto no se hizo (grupo de control), distribuidas en los siete estados piloto en que se inició la aplicación del programa: Guerrero, Hidalgo, Michoacán, Puebla, Querétaro, San Luis Potosí y Veracruz. Incluyó a 24 077 hogares y más de cien mil personas, o sea, aproximadamente 1% de los hogares apoyados.

Skoufias señala que este importante tamaño muestral, podría ser incluso insuficiente para medir el impacto en algunas áreas; lo que de inmediato refuerza las dudas que dejan las evaluaciones de otros programas sociales en que la muestra dista mucho de alcanzar este tamaño, independientemente de los problemas enfrentados para garantizar una selección aleatoria.

Entre los resultados globales obtenidos de esta forma para Progresá, la evaluación realizada destaca, a nivel del diseño del programa, una focalización más clara de lo que sucedía habitualmente en programas similares de América Latina, una importante sinergia debida a la consideración de tres sectores-clave (salud, educación y alimentación), la utilización de transferencias condicionadas a la participación de las familias beneficiadas en escuelas y centros de salud, y el apoyo complementario a los últimos. Concretamente, destaca:

1. Muy buena focalización geográfica en zonas rurales,
2. Bastante precisión en el método de selección de hogares,
3. Fuerte impacto en el ingreso real de las familias,
4. Ausencia de impactos negativos como disminución del tiempo dedicado por los padres de familia al trabajo,
5. Errores administrativos menores como demora en la distribución de las transferencias, errores en listas de beneficiarios, etcétera,
6. Efectos indirectos positivos en las localidades con familias apoyadas debido a la derrama promedio de \$30 000 mensuales adicionales.

A pesar de la rigurosidad del diseño evaluativo, el lenguaje empleado en la presentación de resultados transmite la conciencia de los evaluadores sobre el

---

encuentran diferencias en algunas variables que, aunque en promedio son muy pequeñas, el gran tamaño de la muestra, hace que resulten significativas a niveles convencionales (10% en su análisis).

grado de riesgo implícito en algunos de ellos: “los resultados de la investigación sugieren...”, “Los datos sugieren que Progresita...” o “Parece que Progresita...” (IFPRI, 2002: 3, 5 y 7, respectivamente). Skoufias insiste en su insuficiencia para determinar algunos impactos que serían deseables, como:

1. ¿Cuál es el monto mínimo a transferir como beca para lograr un impacto real en la permanencia de los niños en la escuela y su acceso regular a los centros de salud?
2. Además de la permanencia en la escuela ¿realmente se incrementó la calidad de la educación y el aprendizaje logrado por los estudiantes? La evaluación no encontró indicios de esto último, que constituiría la misión del programa educativo.
3. ¿El programa puede constituirse en una red de seguridad social en momentos de crisis?
4. ¿Cuánto tiempo puede continuar un programa como éste, que requiere de un monto importante de recursos económicos, si no demuestra claramente su eficacia?

Estas fueron algunas de las dudas que quedaron planteadas en las evaluaciones efectuadas por el IFPRI, a las que se agregan diversas condiciones no cumplidas que Skoufias (2003) considera necesarias para garantizar una evaluación exitosa y rigurosa: que la evaluación sea un componente incluido desde el diseño del programa, que tenga objetivos claros, que se cuente con el apoyo del ámbito institucional responsable del mismo, que el diseño de la evaluación ofrezca mediciones creíbles, que cuente con un adecuado marco lógico, presupuesto suficiente y, de preferencia, un diagnóstico previo a la aplicación del programa.

Durante el año 2000 Progresita atendió a 2 478 millones de hogares. Se continuó con la aplicación de las encuestas semestrales a hogares, en dos nuevas rondas (en mayo y noviembre) para completar los tres años de seguimiento, mismo que había dado inicio en marzo de 1998. La novedad fue que, en estas fechas, el grupo de control había sido incorporado al programa, con lo que el diseño experimental original sufrió un cambio: ya no había la contundente posibilidad de comparar resultados entre beneficiarios y no beneficiarios, sino sólo entre dos grupos de beneficiarios, incorporados al programa con menos de

dos años de diferencia (mayo de 1998 y enero de 2000). Las razones del cambio, que metodológicamente dificultaron su evaluación, fueron de carácter ético: imposibilidad de continuar negando al grupo de control su participación en los beneficios generados por el programa, siendo que los recursos eran suficientes para su inclusión.

### **Auditorías de desempeño de Progres a 1999-2000**

En forma paralela al proceso descrito, a partir de 1999, la ASF comenzó a realizar las denominadas “auditorías de desempeño” a los programas federales. Entre éstos incluyó al Progres a: su componente educativo en 1999<sup>7</sup> y su componente de salud en 2000. Los mismos fueron analizados como parte de la responsabilidad que corresponde a la SEP y la Ssa, no a la Sedesol [[www.asf.gob.mx](http://www.asf.gob.mx)].

Los informes de las auditorías de Progres a comienzan por fundamentar las razones que llevaron a la selección del programa, se apoyan en información documental del mismo y realizan un trabajo de campo específico para el cual solicitan el apoyo de otras instituciones (UNAM). También retoman los informes de los estudios realizados por el IFPRI, el BM y la Comisión Económica para América Latina.

Su trabajo utiliza un conjunto de indicadores para el análisis de cada programa, en el que figura:

- Congruencia del objeto social del programa con las prioridades nacionales y sectoriales de mediano plazo (el sexenio en curso),
- eficacia en el cumplimiento de las metas programadas,
- cobertura respecto del universo objeto del programa,
- eficiencia en la operación y aplicación del presupuesto, e
- impacto del programa.

<sup>7</sup> No fue posible acceder a dicho documento debido a que sólo están disponibles en internet los informes de la Cuenta Pública del año 2000 en adelante.

La auditoría de desempeño aplicada a Progresas<sup>8</sup> en el año 2000 realiza observaciones y promueve acciones específicas, entre las que resalta la insistencia en que el programa no había establecido indicadores de impacto que permitieran evaluar el cumplimiento de su objetivo (mejorar el nivel de bienestar y calidad de vida de la población beneficiaria) en su programación anual y de mediano plazo. Entre las recomendaciones de menor magnitud se encuentran: poner en marcha un sistema de quejas y denuncias, realizar campañas de difusión del componente de salud, controlar que el personal médico y paramédico no exija ninguna aportación por la entrega del suplemento alimenticio, entre otras.

### **Evaluación y auditoría de desempeño de Progresas 2001**

En 2001 se pusieron en marcha dos nuevas acciones: la ampliación de la cobertura del programa hacia las zonas urbanas, y la entrega de becas educativas a estudiantes de nivel medio superior.

De esta forma se incorporaron al programa, originalmente diseñado sólo para el medio rural, 600 localidades urbanas y 1.76 millones de nuevas familias, o sea que el número de familias atendidas aumentó 72%. Para el ciclo escolar 2001-2002, según datos de Sedesol, se logró la inscripción de 3.3 millones de becarios en los niveles de primaria, secundaria y educación media superior, cifra superior en 34% al número de niños y jóvenes atendidos en el ciclo escolar 2000-2001. En particular se entregaron becas educativas para continuar estudios de nivel medio superior a 267 733 jóvenes que vivían en condiciones de pobreza extrema.

Por otra parte, para corregir errores de exclusión en el proceso de selección de familias se estableció un mecanismo de densificación de la cobertura en las localidades rurales ya atendidas. Con el fin de atender a los beneficiarios del programa en las ciudades se instalaron 196 módulos que operan bajo un esquema automatizado de atención, basado en el uso de tecnologías informáticas y de transmisión electrónica de datos. Cada módulo fue habilitado con equipo de cómputo y una plataforma de cartografía digitalizada, lo que permitió reducir el

<sup>8</sup> Sorprende que en diversas partes del documento se hable de Oportunidades (aún no creado en el año 2000) en lugar de referirse a Progresas. Consúltense, por ejemplo, las acciones promovidas 00-20G00-7-373-07-001, 00-20G00-7-373-07-002 y 00-12100-7-373-07-008.

tiempo desde el levantamiento de la información socioeconómica de las familias hasta su incorporación.

El programa le solicitó a una institución nacional que propusiera los instrumentos de línea basal urbana en pequeñas localidades. Dicha línea fue diseñada entre mayo y septiembre con grupo de tratamiento y de control y se inició el trabajo de campo. A fines de 2001, al realizarse las modificaciones ya mencionadas al programa, se abandonó el seguimiento de los paneles y se decidió la ampliación del programa a ciudades de mayor tamaño, sin respetar el grupo de control diseñado.

Estrictamente, la evaluación de la gestión fiscal 2001, a diferencia de las anteriores que duraron dos años y medio, fue realizada en un mes. La misma incluyó: una evaluación del funcionamiento del programa a nivel estatal, basada en el análisis de los indicadores de seguimiento, gestión y evaluación de 1998 al 2001; y, un análisis de los impactos del programa basados en las rondas recientes de la Encuesta de Evaluación en Localidades (Encel), no utilizadas en las evaluaciones previas. O sea que no incluyó levantamiento específico de información en campo. Su costo fue de \$16.13 millones.

La evaluación identifica problemas relevantes en el mencionado sistema de indicadores (respecto de la continuidad de los incluidos, de su desglose, y de su formato de presentación, algunas redundancias y una clasificación no es siempre clara, algunos indicadores cuya relevancia para la evaluación del Programa es limitada, y otros más relevantes que están ausentes) y dificultades para hacer una evaluación a nivel estatal con datos agregados, por lo que señala que sus resultados son indicativos y ameritan estudios en mayor profundidad. Entre éstos resaltan:

- Una distribución de las transferencias monetarias por estado congruente con sus niveles de pobreza monetaria y las necesidades relevantes de salud, lo cual no resulta igualmente claro en materia educativa.
- Avances notables en el cumplimiento de las condiciones de corresponsabilidad que impone el programa a sus beneficiarios en materia de educación y salud.
- Avances claros en relación con las tasas de deserción escolar y de desnutrición infantil.

- Persistencia de altas tasas de repetición de los beneficiarios y de una brecha importante entre éstos y los promedios estatales en materia educativa [[www.progresa.gob.mx](http://www.progresa.gob.mx)].

Por su parte, las nuevas rondas de Encel dieron lugar a dos estudios de medición de impacto: un análisis del comportamiento del grupo de control, al ser incorporado al programa; y una evaluación de algunos resultados de mediano plazo (solo tres años) en los jóvenes beneficiarios del mismo. Los últimos incluyeron el cálculo de la probabilidad de que los estudiantes becados por Progresía en el nivel básico continuaran con su educación media superior (aún no se otorgaban becas para ese nivel), o, si se incorporaban al mercado de trabajo, obtuvieran mejores empleos y remuneraciones, así como su relación con la migración, y la propensión de los niños a enfermar.

De los resultados obtenidos en los anteriores, se destaca:

- Progresía benefició al 40% de las familias en las zonas rurales.
- El grupo de control se comportó en forma muy similar al de tratamiento, al momento de su inclusión al Progresía, en aquellas variables en que se puede identificar un efecto inmediato, como inscripción escolar, visitas a centros de salud y gasto de los hogares.
- Se confirmaron también efectos en los años de escolaridad de los niños y niñas apoyados por el programa: 0.15 y 0.38 años más, respectivamente, que los niños que no recibieron becas.
- Los datos “sugieren” un pequeño efecto en la inscripción de los exbecarios a la educación media superior, que se esperaba se incrementara con el tiempo, lo que se reforzaría por un nuevo elemento: el otorgamiento de becas de Progresía a este nivel a partir de 2001.
- La información disponible “no sugiere” que los jóvenes de las localidades donde opera Progresía tengan mayor probabilidad de migrar fuera de sus comunidades, incluso cuando han dejado de ser elegibles para la obtención de becas de estudio.
- Resulta muy prematuro evaluar los efectos del programa en las oportunidades de empleo e ingresos de los exbecarios, ya que la mayoría aún continúa en la escuela, siendo los que la abandonaron quienes no lograron mejorar su desempeño a pesar de la beca [[www.progresa.gob.mx](http://www.progresa.gob.mx)].

Como ya se comentó, metodológicamente hablando, el diseño evaluativo se debilitó con la inclusión del grupo de control al programa, por lo que se recomendó la utilización de una estrategia no experimental de estimación<sup>9</sup> para la medición de impactos del programa en el largo plazo.

Ese mismo año se presentaron resultados de una evaluación cualitativa realizada en zonas semiurbanas. Con estos dos trabajos, se cubrió la evaluación anual de 2001, si bien los resultados se basaban en información recogida mediante las Encel de mayo y noviembre del 2000.<sup>10</sup>

Al igual que el año anterior, la ASF llevó a cabo la auditoría de desempeño de programas a cargo de la Sedesol e incluyó nuevamente al Progresá en su componente alimenticio (dentro del sector salud).

En esta ocasión su informe reconoce que el programa cuenta con los indicadores y evaluaciones necesarias para medir el cumplimiento de sus objetivos y metas, así como sus procesos operativos. Recomienda dar cabal cumplimiento a las normas que rigen la rendición de cuentas y reforzar mecanismos de control y supervisión en las unidades médicas relativas al cumplimiento de las acciones que deben realizar las familias beneficiarias como parte de su corresponsabilidad.

Cabe destacar también que Progresá fue uno de los pocos programas sociales evaluado por organizaciones no gubernamentales. En 1999, y como parte del seguimiento a los acuerdos de Beijing, la campaña “El Banco Mundial en la mira de las mujeres” analizó un conjunto de programas de América Latina que, en el caso de México, incluyó al Programa de ampliación de cobertura en salud (PAC) y al Progresá (Frade, 2000). La evaluación relativa a este último se diseñó con base en una muestra de 309 beneficiarias y 123 no beneficiarias, todas pertenecientes a organizaciones sociales y distribuidas en los estados de Chiapas, Campeche, Guanajuato, Jalisco, Oaxaca, Puebla, Yucatán y Quintana Roo, a las que se aplicó una encuesta durante el segundo semestre de 1999.

<sup>9</sup> Sugiere la estrategia de “matching”, emparejamiento o selección de pares, véase Heckman, J.J., R. Lalonde y J. Smith (1999), “The economics and econometrics of the active labor market program”, en *Handbook of labor economics*, vol. 3A, ed. O. Ashenfelter and D. Card Amsterdam, The Netherlands, Nort Holland.

<sup>10</sup> En realidad las evaluaciones de Progresá-Oportunidades nunca correspondían estrictamente al año calendario, como ha sucedido con los otros programas. Incluso la evaluación 1998-2000, utilizó las bases de datos disponibles hasta finales de 1999.

Además se realizaron 83 entrevistas a maestros, promotores del programa, personal de centros de salud, funcionarios federales, estatales y municipales, así como organizaciones sociales.

Entre sus resultados reconoce la asignación de recursos gubernamentales, los costos de operación bajos del programa (aunque basándose en el trabajo gratuito de las mujeres) y el diseño de un componente de género. Sin embargo, su juicio es muy negativo en cuanto a la ausencia de participación corresponsable de las beneficiarias en la toma de decisiones, seguimiento y evaluación del programa, desigualdad y falta de criterios para la asignación de apoyos, exclusión de las zonas rurales de mayor pobreza por falta de infraestructura educativa y de salud, cobro de cuotas importantes por servicios supuestamente gratuitos, no consideración del nivel de desempleo en el país y de las escasas oportunidades existentes en el medio rural, trato indigno de las beneficiarias como indigentes, generación de divisiones entre la población, constitución de un mecanismo de control social, falta de adecuada conceptualización de la perspectiva de género, ignorancia de problemas relevantes en ese campo, entre otros. Resalta la forma en que se refieren al programa algunas entrevistadas del estado de Puebla, “dinero del diablo”, porque “los recursos no les alcanzan cuando llegan, ‘no rinden’, pero, además, no se los ganaron” (Dora Ramos y compañeras, en Frade, 2000:86). Como consecuencia de los múltiples problemas identificados, el informe concluye que Progresá no permite superar la pobreza ni mejora la posición de las mujeres.

Finalmente, y con una opinión totalmente contrapuesta a la previa, el propio BM (2004) considera que México ha sido pionero por la aplicación de evaluaciones de impacto de Progresá, las que permitieron al gobierno mexicano contar con evidencias de su éxito y fomentaron la creación de programas similares en otros países de América Latina (Honduras y Nicaragua con diseños experimentales y Colombia y Brasil con cuasiexperimentos).

### **Evaluación de programas federales a partir de 2001**

La evaluación del desarrollo social a nivel federal presenta un claro parteaguas con la creación del Coneval en 2006, por lo que esta presentación se divide en dos etapas, antes y después del mencionado evento.

*La etapa 2001-2005*

Como se ha visto, el proceso de evaluación de programas gubernamentales comenzó a instrumentarse de manera sistemática a partir de 1997 con la creación del Progres a y se generalizó en forma obligatoria para todos los programas gubernamentales sujetos a RO, con base en la decisión incluida en el Presupuesto de Egresos de la Federación (PEF) aprobado por la Cámara de Diputados en el año 2000; en particular, los dirigidos a la reducción de la pobreza que se encontraban a cargo de la Secretaría de Desarrollo Social.

En 2001 no se realizaron evaluaciones externas a todos los programas y algunas se efectuaron sólo a un semestre del mismo, por lo que las efectuadas a la gestión 2002 constituyeron la primera generación completa.

Durante las gestiones 2001-2005, las evaluaciones externas fueron organizadas fundamentalmente por medio de la Subsecretaría de Prospectiva, Planeación y Evaluación de la Sedesol y, a partir de 2002, se guiaron por el “Acuerdo por el que se expiden los requisitos mínimos que las dependencias y entidades de la Administración Pública Federal deberán considerar para seleccionar a las instituciones académicas y de investigación u organismos especializados, de carácter nacional o internacional, interesados en realizar la evaluación de resultados de los programas gubernamentales sujetos a RO en el ejercicio fiscal 2002” firmado por la Sedesol y la entonces Secodam (en adelante el Acuerdo de 2002), concentrándose en el análisis de resultados, con pequeñas variantes como las señaladas por Conde (2007): incluir el impacto sobre el bienestar, la equidad, la igualdad y la no discriminación de las mujeres en 2002 a 2004, o informar sobre los recursos entregados a los beneficiarios en 2003. En 2005 se solicitó realizar un metaanálisis 2002-2005, lo que implicaba evaluar la gestión 2005 y luego proceder a la comparación de los resultados encontrados.

En 2002 existieron 31 programas evaluados, mientras que en 2005 fueron sólo 16, de los cuales 11 formaban parte del grupo inicial (salvo cambios de nombre). Se identificaron como programas nuevos: Atención a los adultos mayores en zonas rurales, Hábitat, Adquisición de leche nacional, Apoyo alimentario, y Vivienda rural (Cardozo, 2009).

Cardozo (2006a) caracterizó la situación de la evaluación en 2002-2003 por:

- Exigencias y expectativas poco realistas debidas a: 1) decisión de la Cámara de Diputados de evaluar todos los programas sujetos a RO, todos los años y en todos sus aspectos, 2) expectativas de Sedesol respecto de la posibilidad y conveniencia de aplicar estudios experimentales, y 3) limitado desarrollo de una cultura de la evaluación que valorara su contribución, cambiara actitudes y se tradujera en nuevas prácticas que integren la evaluación a la planeación y la operación.
- En materia del marco normativo aplicable, la Sedesol apoyó su trabajo en el artículo 134 de la Constitución Política de los Estados Unidos Mexicanos; 26 fracción II, 40, 41, 42 y 43 de la Ley de Adquisiciones, Arrendamientos y Servicios del Sector Público (en adelante, Ley de Adquisiciones), el artículo 32 de la Ley Orgánica de la Administración Pública Federal, los sucesivos decretos de PEF, y el Acuerdo de 2002, fundamentalmente. Entre los aspectos problemáticos de dicho marco resalta: 1) la primera de las citadas leyes está pensada para comprar productos y, a lo sumo, algunos servicios, pero se adapta poco a la contratación de servicios de consultoría para la evaluación, lo que, en la práctica, generó dudas como la posibilidad de otorgar contratos por adjudicación directa en lugar de proceder a la realización de licitaciones públicas; y 2) el Acuerdo, las bases y los TR incurrieron en múltiples imprecisiones y ausencias por falta de experiencia en su redacción.
- En relación con los objetivos y tipos de evaluación, se definió que el primero se dirigiría a institucionalizar un Sistema de Evaluación que incluiría siete componentes: 1) evaluación externa de corto plazo, 2) evaluación externa de largo plazo, 3) evaluación interna de diseño e impacto, 4) evaluación interna de existencia de obras y sustentabilidad, 5) evaluación de la política social; 6) evaluación cualitativa, y 7) fomento a la formación de capital humano en temas de evaluación. Evidentemente, el componente de evaluación cualitativa, igual que el de evaluación cuantitativa no mencionado, podría ser utilizado en cualquiera de los primeros cinco tipos de evaluación. En función de nuestro marco teórico, dichas evaluaciones podían clasificarse como:
  - Parcialmente formativas, ya que realizan su análisis con fines de mejoramiento de los programas pero, en la práctica, son poco utilizadas.

- *Ex-post*, debido a que su valoración se realiza anualmente y después de acabada dicha gestión.
- Externas, porque los evaluadores carecen de vinculación directa con los operadores de los programas, funcionarios y congresistas.
- De proceso, resultados e impactos, ya que la evaluación del diseño no fue solicitada en el Acuerdo y los Términos de referencia.
- En materia de organización formal, participaban en el proceso múltiples instancias:
  - La Cámara de Diputados, a propuesta del Ejecutivo, dio origen al proceso de evaluación a partir del año 2001.
  - La decisión anterior afectó a todas las Secretarías que desarrollaban programas con recursos fiscales sujetos a RO, en particular, a la Secretaría de Desarrollo Social.
  - En 2002, la Sedesol contó con 31 programas, 13 operados en forma directa por su Dirección General de Políticas Sociales y otros 18 sectorizados en empresas y organismos desconcentrados.
  - Cada uno de los anteriores contaba con una oficina encargada de efectuar las licitaciones, contratos, control de la entrega de informes convenidos, relacionarse con las instituciones y los equipos de investigadores que realizaban las evaluaciones, enviar los informes finales a la Sedesol y autorizar el pago al evaluador.
  - Las universidades y centros de investigación contratados para realizar los trabajos.
  - Por último, la Sedesol realizaba comentarios a los evaluadores y entregaba los informes a la Cámara de Diputados, SHCP y Secretaría de Función Pública.

En la práctica, la Cámara de Diputados no realizaba la función prevista en la decisión formal, y la SHCP y la SFP cumplían un papel muy restringido, por lo cual el proceso quedaba a cargo de la Sedesol, como principal destinataria real de los resultados de las evaluaciones, sin realizar una amplia difusión de sus resultados.

- En torno a los recursos financieros y de información disponibles para la realización de los trabajos, éstos fueron en general, insuficientes; sin embargo,

Oportunidades llegó a disponer de presupuestos excesivos para la calidad de los trabajos realizados, lo que mostró la falta de criterios explícitos para definir techos financieros. Adicionalmente, la decisión de que las propias instancias que operaban los programas fueran quienes contrataran y pagaran a los evaluadores provocó serios riesgos para la necesaria autonomía de los últimos.

Por su parte, los recursos humanos mostraron una oferta restringida de capacidades especializadas en el tema, debido a falta de experiencia, agudizada por la ausencia de programas universitarios de formación y capacitación, de disposición de guías de evaluación, y de asociaciones de profesionales especializados en la materia.

Un evento de relevancia, posterior al análisis anterior, lo constituye la expedición de LGDS en enero de 2004, misma que preveía plazos que no se cumplieron para la publicación de su reglamento y la conformación de un Consejo de Evaluación.

### *El periodo iniciado en 2006*

En 2006 se regresó a dirigir las evaluaciones al diseño, cobertura y focalización, operación, resultados y percepción; aunque Conde (2007) expresa que la Sedesol introdujo modificaciones a su contenido, apartándose del Acuerdo de 2002, debido a que solicitó tomar en cuenta algunos de los aspectos ya propuestos en nuestro trabajo anterior, como la evaluación del diseño del programa, e introdujo el meta-análisis 2002-2005.

En un nuevo contexto normativo que se revisará de inmediato, en 2007 se hicieron evaluaciones de diseño referidas a las características del programa, su contribución a los objetivos estratégicos, su matriz de indicadores, las RO, así como las coincidencias, complementariedades o duplicidades con otros programas. Además se iniciaron evaluaciones de consistencia y resultados, que finalizaron en 2008.<sup>11</sup> Éstas implicaron un diagnóstico sobre la capacidad institucional, organizacional y de gestión de los programas, y se basaron

<sup>11</sup> Este proceso duró aproximadamente nueve meses, más que los predecesores.

exclusivamente en información de gabinete. Por otro lado, Coneval solicitó directamente cinco evaluaciones de programas: Abasto social de leche, Apoyo alimentario, Abasto rural, Hábitat y Oportunidades. En 2008, se realizaron las evaluaciones de programas contratadas por ellos mismos (Cardozo, 2009) y Coneval solicitó la Evaluación de la política de desarrollo social en México, que incluye un diagnóstico de la misma a partir de la década de 1990. En 2009 y 2010 se realizaron evaluaciones de desempeño de programas y, en el último de los años señalados, también una Evaluación integral del desempeño de los programas federales.

La nueva caracterización del periodo es retomada de Cardozo (2009):

- En torno a las exigencias y expectativas, aunque la LGDS continúa pretendiendo evaluar todos los programas sujetos a RO, todos los años y en todos sus aspectos, en la práctica, Coneval buscó la forma de evitar un compromiso que no resulta realísticamente factible.
- En relación con el marco normativo, a inicios del 2006, se produjo la expedición del Reglamento de la LGDS, que provocó serias discrepancias motivadas por las modificaciones que introdujo a la Ley correspondiente, lo que estuvo a punto de transformarse en una controversia constitucional (Boltvinik, varios 2005 y 2006).

En marzo de 2007, la SHCP presentó a la Cámara de Diputados, en cumplimiento de lo previsto en la Ley Federal de Presupuesto y Responsabilidad Hacendaria, su Propuesta de SED. La Comisión de Presupuesto y Cuenta Pública (CPyCP) reconoció su contribución al diseño de nuevos mecanismos para la asignación de presupuesto, pero estuvo en desacuerdo porque no se incluían los mecanismos de participación de la Cámara de Diputados, a partir de sus Comisiones Ordinarias, coordinadas por la CPyCP, se confería al Ejecutivo Federal un papel de “juez y parte”, daba facultades exclusivas a la SHCP y al Coneval en detrimento de las correspondientes a la SFP y se corría el riesgo de priorizar los objetivos de control del gasto sobre los del desempeño. Por otra parte, la Comisión de Presupuesto consideró importante rescatar la experiencia iniciada en 2001 por la ASF en la realización de auditorías de desempeño, improcedente la propuesta de transformar al Coneval en un Consejo Nacional de Evaluación de las Políticas Públicas, y refrendó las facultades evaluatorias de sus Comisiones Ordinarias (CEDRSSA, 2008).

También en marzo de 2007, el Coneval publicó en el *Diario Oficial de la Federación*, conjuntamente con la SHCP y SFP, los “Lineamientos generales para la evaluación de programas federales de la Administración Pública Federal” (en adelante, los lineamientos), en sustitución del Acuerdo de 2002. Su objetivo fue regular la evaluación de los programas federales, la elaboración de la matriz de indicadores y de los sistemas de monitoreo, y la definición de los objetivos estratégicos de las dependencias y entidades de la administración pública federal.

Se mantuvo como posibilidad la criticada exigencia previa de que las evaluaciones externas fueran pagadas con cargo al presupuesto de la dependencia o entidad responsable del programa, cuestión que casi todos los entrevistados están de acuerdo en que es limitativa de la libertad del investigador. Se agregó la opción de que lo haga el Coneval, en el ámbito de su competencia y cuando éste así lo determine.

También se estableció la obligación de: 1) publicar un Programa Anual de Evaluaciones en el que se dé a conocer los programas que serán evaluados y el tipo de evaluación a realizar (capítulo I, artículo décimo séptimo), 2) dar seguimiento a las recomendaciones efectuadas por los evaluadores con base en el convenio de compromisos de mejoramiento de la gestión para resultados (capítulo VII, artículo vigésimo quinto) y 3) publicar los informes de las evaluaciones en las páginas electrónicas de las dependencias y entidades responsables de los programas y en el portal del SED (capítulo VII, artículos vigésimo sexto y vigésimo séptimo).

Una innovación importante en términos de transparencia es la información complementaria, antes muy difícil de obtener, que también debe publicarse con datos generales como: nombre del coordinador de la evaluación externa y los miembros principales de su equipo, unidad administrativa que da seguimiento a la evaluación, forma de contratación, tipo de evaluación contratada, bases de datos, instrumentos de recolección de información, nota metodológica, resumen ejecutivo y costo total de la evaluación, incluyendo su fuente de financiamiento (capítulo VII, artículo vigésimo octavo).

Finalmente, el capítulo III establece los requisitos mínimos de elegibilidad de los evaluadores externos y los procedimientos de contratación que deben mantenerse apegados a la Ley de Adquisiciones.

Con este nuevo marco, que mejoró sensiblemente el tipificado en el Acuerdo de 2002, quedó pendiente uno de los problemas más serios, la falta de carácter vinculatorio de las recomendaciones que surgen de las evaluaciones realizadas. Con la intención de aumentar su utilización en el rediseño de políticas y programas, en octubre de 2007, el Coneval, conjuntamente con la SHCP y la SFP, emitió el “Mecanismo para el seguimiento a los aspectos susceptibles de mejora derivados de los informes y evaluaciones externas a programas federales 2008”.

También el tema de la selección, contratación y pagos, sufrió algunos cambios menores<sup>12</sup> debido a las modificaciones producidas en la Ley de Adquisiciones, mismos que podrían facilitar los procesos de selección de evaluadores, pero que también podrían provocar contrataciones discrecionales, no fundamentadas en razones técnicas: 1) la fracción III del artículo 36 que permite considerar para la adjudicación no sólo el precio más bajo sino la calidad de la propuesta;<sup>13</sup> y 2) la fracción X del artículo 41 que permite efectuar contrataciones, sin sujetarse al procedimiento de licitación pública, a partir de los procedimientos de invitación a cuando menos tres personas o de adjudicación directa cuando se trate de servicios de consultorías, asesorías, estudios, investigaciones o capacitación, debiendo aplicar el procedimiento de invitación a cuando menos tres instancias, entre las que se incluirán instituciones de educación superior y centros de investigación.

- En cuanto a los objetivos y tipos de evaluación, éstos quedaron establecidos en los nuevos lineamientos: “el análisis sistemático y objetivo de los programas federales y que tiene como finalidad determinar la pertinencia y el logro de sus objetivos y metas, así como su eficiencia, eficacia, calidad, resultados, impactos y sostenibilidad” (capítulo I, artículo tercero, numeral IV). Esta definición deja

<sup>12</sup> Nos referimos a las reformas y/o adiciones del 13 de junio de 2003, 7 de julio de 2005, 21 de agosto de 2006, 20 de febrero de 2007, 5 de septiembre de 2007 y 1 de octubre de 2007.

<sup>13</sup> Tratándose de servicios, podrá utilizarse el mecanismo de puntos y porcentajes para evaluar las propuestas, en el que el rubro relativo al precio tendrá un valor porcentual del 50%, indicando en las bases la ponderación que corresponderá a cada uno de los demás rubros que serán considerados en la evaluación, de acuerdo con los lineamientos que para tal efecto emita la Secretaría de la Función Pública. Asimismo, cuando sea necesario, en el caso de servicios se solicitará el desglose de precios unitarios, precisando de qué manera será utilizado éste.

fuera la parte interpretativa y valorativa propia de la evaluación y en ella parece repetirse el criterio de eficacia, que no es otra cosa más que el logro de objetivos y metas, mientras que el gran ausente es el relativo a la equidad.

Las evaluaciones fueron clasificadas en dos tipos: estratégicas y de programas federales (capítulo I, artículo décimo sexto). Las primeras se aplican a las estrategias, políticas e instituciones relacionadas con un programa o conjunto de programas; mientras que dentro de las segundas se identifican: evaluaciones de consistencia y resultados, de indicadores, de procesos, de impacto, y evaluaciones específicas. Se agrega que los principales contenidos de las evaluaciones de consistencia y resultados se ajustarán a TR genéricos (capítulo II, artículo trigésimo) e incluirían (capítulo II, artículo décimo octavo): diseño del programa, planeación estratégica, cobertura y focalización, operación, percepción de la población objetivo, y resultados. Nuevamente, los indicadores no constituyen un tipo específico, sino que podrían utilizarse en cualquiera de las demás. Posteriormente, se definieron 34 preguntas binarias para todos los programas que evaluaran su diseño y 100 en el caso de las relacionadas con consistencia y resultados.

- En cuanto a la organización del sistema de evaluación, el mismo se mantiene en forma similar al pasado, salvo la sustitución de la responsabilidad a cargo de la Sedesol por las nuevas atribuciones del Coneval. La Cámara de Diputados continúa sin realizar su función en el proceso. Por su parte, la SHCP y la SFP se involucran más con la propuesta del SED y los compromisos asumidos para dar seguimiento a las recomendaciones incluidas en las evaluaciones. Se han ampliado de manera importante los mecanismos de difusión por internet, no así otras vías para la información a la ciudadanía en general y en especial a los beneficiarios.
- En materia de recursos, las ocho evaluaciones de la gestión de 2005 encargadas por Sedesol (la mitad de las realizadas) tuvieron un costo de \$400 225.00, en promedio, si bien no incluyeron las evaluaciones más costosas (Oportunidades, por ejemplo). En cambio, en 2007, las cinco evaluaciones de diseño, con base exclusivamente en trabajo de gabinete, costaron en promedio \$262 057.41; por su parte, las 16 de consistencia y resultados realizadas en 2008 (que incluyeron programas importantes como Oportunidades, Hábitat, Empleo temporal, Abasto social de leche, etcétera), también de gabinete, insumieron

en promedio \$554 060.60.<sup>14</sup> Si comparamos estos datos con el costo promedio de las 31 evaluaciones realizadas en 2002 de \$1 695 000.00 (Cardozo, 2006b), en total, se están dedicando menos recursos a la evaluación que en el pasado (\$8 864 969.55 en 2008 frente a \$20 294 004.70 en 2002) debido a la anterior aplicación de trabajo de campo. Sin embargo, las autoridades de Coneval consideraban que los recursos habrían aumentado, al agregarse los aportados por dicho Consejo.

A nivel de los recursos humanos también hay acuerdo en que éstos han venido perfeccionándose en la práctica,<sup>15</sup> no por la realización de estudios formales, que son prácticamente inexistentes en el país. Algunos equipos se han consolidado a partir de su experiencia previa en la realización de encuestas y muestreos, otros en forma autodidacta o tomando cursos como los ofrecidos por el BID.<sup>16</sup> También se menciona que la mayoría de los evaluadores se han especializado en evaluaciones de procesos, gestión y percepción, pero son muy pocos los que lo han hecho en medición de impactos.

En relación con los insumos de información se menciona que aunque su disposición y sistematización ha ido mejorando en los últimos años, su confiabilidad, suficiencia y oportunidad varía sustancialmente de un programa a otro. También hay problemas con la introducción de cambios en los campos de las bases de datos o en la apertura programática que dificultan o imposibilitan la comparación entre diferentes ejercicios fiscales. Las necesarias depuraciones disminuyen el tiempo disponible para la realización de las evaluaciones, por lo que se considera que aún se requiere mejorar mucho más (Cardozo, 2009).

Finalmente, con base en las evaluaciones realizadas en 2007, se percibía una tendencia a: generalizar el empleo del monitoreo basado en el Marco Lógico, asignar contratos de evaluación sin licitación previa (salvo las directamente solicitadas por Coneval), realizar evaluaciones con mayores limitaciones de

<sup>14</sup> Información entregada por Sedesol o publicada en la página electrónica de Coneval.

<sup>15</sup> El funcionario entrevistado en SFP insistió en indicar que la mejoría es muy pequeña.

<sup>16</sup> Sin embargo hay situaciones inaceptables: por ejemplo, en 2006 de 15 evaluaciones realizadas sólo seis especificaron el tamaño de sus muestras y una de las encuestas se aplicó telefónicamente.

la libertad del evaluador para definir su metodología (respuesta de un listado predeterminado de preguntas, ausencia de trabajo de campo) y, con una correspondiente y drástica reducción de presupuestos.

### **Evaluación de políticas y programas del Gobierno del Distrito Federal**

En 2008, el gobierno local del Distrito Federal instaló Evalúa DF, compuesto por el secretario de Desarrollo Social del DF, y seis consejeros ciudadanos seleccionados en convocatoria pública por la Comisión Interinstitucional de Desarrollo Social del Distrito Federal que, al igual que en el caso de Coneval, constituye un avance hacia la participación ciudadana en la evaluación, aunque todavía de carácter restringido. Evalúa DF había sido creado el año anterior por decreto de su jefe de Gobierno, ante la necesidad de instrumentar un proceso, casi inexistente en su administración, que contribuyera a mejorar la gestión de sus programas.

Este Consejo local ha organizado un sistema de evaluación sustantivamente diferente al federal, cuyo contraste se describe a continuación. Evalúa DF decidió discutir, analizar y capacitar a los servidores públicos en el tema, en forma paralela al inicio de sus trabajos, debido al reconocimiento de la limitada aceptación cultural que la evaluación tiene entre los funcionarios mexicanos, el relativo nivel de institucionalización de esta función y de profesionalización de quienes la realizan; proceso que no se realizó a nivel federal. Espera que el evaluador recoja y juzgue evidencias, pero también que actúe como facilitador de un proceso de aprendizaje, lo que a mediano plazo implica un cambio en su rol y en las características requeridas para cumplirlo (interdisciplinarietà, habilidades de negociación, didácticas, etcétera), aspectos resaltados en la experiencia francesa (Monnier, 1995).

Resaltar el papel positivo de la evaluación, ligado al aprendizaje de los servidores públicos y de la ciudadanía, en contraste con el temido enfoque punitivo aplicado por el GF durante el siglo pasado, ha sido uno de los objetivos importantes durante la capacitación señalada. También se han identificado las limitaciones de este proceso, de manera que las expectativas generadas sean realistas.

En Evalúa DF se pretende retomar la experiencia de países como Francia, Gran Bretaña o Canadá para la elaboración de una guía de evaluación apropiada

a nuestro contexto; ya que, a nivel federal, la falta de capacitación de los funcionarios en el tema se agravó debido a la ausencia de un manual básico o guía de evaluación que apoyara específicamente el nuevo trabajo a realizar.

La evaluación anual de todos los programas gubernamentales federales sujetos a RO, en todos sus aspectos (recursos, procesos, resultados e impactos) es considerada demasiado ambiciosa y con un alto riesgo de transformar a la evaluación en un procedimiento superficial y “rutinizado” (como el que vivió Canadá a inicios de la década de 1990), que busca cumplir una disposición legal sin hacer una contribución real al mejoramiento de los programas. Para evitar dicho riesgo, en 2008, Evalúa DF solicitó la evaluación externa de diseño y operación (postergando la medición de resultados e impactos) de sólo 13 programas, seleccionados con base en una combinación de diversos criterios (presupuesto del programa, visibilidad y sensibilidad ante la población, programas “centrales” y delegacionales, entre otros). En 2009 se optó por solicitar la evaluación de la formulación e implementación de cuatro políticas sociales (no sólo de programas), nivel en que la experiencia internacional es casi inexistente. En 2010 se contrató nuevamente la evaluación de otras tres políticas y una metaevaluación; y, en 2011, se está llevando a cabo una evaluación participativa de procesos, resultados e impacto.

Para evitar los procedimientos federales de contratación de los evaluadores centrados en la invitación a tres o más instituciones y, en casos extremos, la asignación directa, Evalúa DF comenzó por hacer una amplia convocatoria para conformar su padrón (el que se mantiene abierto en forma continua), seleccionando a los interesados que demostraron contar con conocimientos y experiencia en la materia. Posteriormente, se invitó a todos los que quedaron inscritos en el anterior a participar, a partir de concurso, en la realización de trabajos. En su selección se tomó en cuenta no sólo el compromiso de cumplimiento de lo solicitado por los TR, sino también la calidad que puede ofrecer el equipo evaluador y su solvencia ética.

Adicionalmente, la experiencia federal muestra equipos de evaluadores que tienden a concentrar varias evaluaciones al mismo tiempo como forma de incrementar sus ingresos, lo que puede reducir la calidad del trabajo. Por ello Evalúa DF limitó a dos el número de evaluaciones que un mismo equipo podría realizar, aunque hasta el presente ningún equipo ha evaluado simultáneamente más de un programa o política.

Como ya se mencionó, un aspecto importante para garantizar lo más posible la necesaria independencia del evaluador es evitar que su pago sea realizado con recursos del mismo programa que está evaluando, problema que se presenta en la experiencia federal. En contraste, Evalúa DF conforma un fondo específico que distribuye los recursos disponibles de manera autónoma y en función de la complejidad del trabajo a realizar. Además, los montos a pagar son transparentados desde la convocatoria y no quedan sujetos a competencia, como sucede en la experiencia federal.

Por otra parte, Evalúa DF interviene a lo largo de todo el proceso de evaluación, en particular, poniendo en práctica procedimientos de acompañamiento del evaluador para superar ágilmente los problemas a que éste se enfrenta. De manera muy especial, en el caso de la contratación de evaluaciones externas (en forma similar al Consejo Científico de la Evaluación en Francia), Evalúa DF opina sobre la propuesta técnica (metodología, fuentes de información, mecanismos de participación, contenido tentativo del informe, cronograma de trabajo, etcétera) que los equipos postulantes presentan y, una vez finalizado el trabajo, elabora un dictamen sobre la calidad del informe presentado: sus datos, métodos, conclusiones y recomendaciones.

Con base en el anterior, Evalúa DF redacta sus propias recomendaciones, mismas que tienen carácter de vinculatorio, diferencia sustantiva con el procedimiento federal, que se espera permita la efectiva utilización de los resultados encontrados para el mejoramiento del diseño, eficacia, eficiencia, equidad y calidad de los programas en curso. Además, este Comité ha diseñado un sistema de seguimiento de la aplicación de dichas recomendaciones y del impacto que van logrando en el programa, de manera de garantizar una verdadera gestión por resultados. En la primera ronda de evaluaciones, se emitieron 121 recomendaciones, de las cuales 85 fueron aceptadas total o parcialmente por los programas evaluados, y con éstas se inició el sistema de seguimiento.

Frente al casi único método de difusión de los informes finales empleado por Coneval (páginas de internet), Evalúa DF considera importante divulgar ampliamente (como lo hace Costa Rica, por ejemplo) una síntesis de los resultados de las evaluaciones en periódicos y revistas de amplia circulación, para fomentar el debate ciudadano. Actualmente, está analizando la posibilidad de publicar en forma de libro algunas evaluaciones seleccionadas.

También en materia de enfoques y métodos de evaluación, hay diferencias sustantivas entre Evalúa DF y Coneval: puesto que los estudios experimentales y cuasiexperimentales requieren la selección de un grupo al que se le aplica el programa y otro de control, con las mismas características, pero que no recibe los bienes, servicios o transferencias entregados por el programa, y a que la política del GDF es prioritariamente universal y basada en derechos, no es posible identificar un grupo excluido del programa, salvo que éste se autoexcluya, en cuyo caso ya no es igual al compuesto por quienes deciden obtener el apoyo del mismo. Por esta razón, por sus altos costos y, muy fundamentalmente, por los múltiples problemas éticos que implica, Evalúa DF se propuso utilizarlos sólo en casos excepcionales.

Por otra parte, Evalúa DF aparece menos preocupado por la objetividad de las mediciones, más abierto a la utilización de métodos tanto cuantitativos como cualitativos, más interesado en la participación de las organizaciones sociales y los ciudadanos en general, y en la conformación de equipos multidisciplinarios donde el peso relativo de cada disciplina depende del programa evaluado.

Este Comité considera también que cada evaluación debe construirse en función del tipo de programa a evaluar (de transferencias, de servicios, de construcción de infraestructura), del área en que se trabaje (salud, educación, vivienda, transporte, etcétera), de la etapa del proceso en que se han detectado problemas (diseño, operación, resultados, impactos), y otros criterios que resulten relevantes en el caso particular.

Asimismo, estima que el evaluador experimentado debe gozar de libertad para definir los métodos con los cuales cumplir con los objetivos de la evaluación solicitada. Sin embargo, al planear la primera ronda de evaluaciones en 2008, ante la limitada información disponible sobre la gestión de los programas, la urgencia por iniciar los procesos, y la demora en la elaboración de una guía de evaluación, Evalúa DF decidió elaborar un amplio cuestionario, casi uniforme, para la evaluación de diseño y operación solicitada en esa oportunidad. Este problema fue corregido a partir de 2009, en que se diseñaron TR *ad hoc* para cada evaluación, que incluyen el listado de asuntos a revisar, pero ya no predefinen cuestionario alguno.

Por último, si bien a nivel federal se ha mencionado el interés de realizar una metaevaluación, con asesoría de un organismo internacional, que analice, valore y contribuya a mejorar la calidad de las evaluaciones realizadas, no se tiene

conocimiento de que la misma haya sido realizada. Evalúa DF, considerando la importancia de reflexionar sobre lo realizado, llevó a cabo una en 2009 y contrató otra en 2010.

### **Comentarios finales**

Los documentos revisados permiten aquilatar la insuficiente información y disponibilidad de conocimientos especializados de evaluación de programas sociales a inicios de la década de 1990 y el importante avance logrado hasta el año 2000, gracias a las evaluaciones realizadas a Progresá por el IFPRI y las auditorías efectuadas por la Auditoría Superior de la Federación.

Con base en estos antecedentes, la Cámara de Diputados estableció un parteaguas en la situación al decidir la ampliación del proceso de evaluación externa a la gestión de 2001 de todos los programas que reciben recursos fiscales y se hallan sujetos a RO. El segundo evento relevante lo constituye la creación del Coneval en 2006. El capítulo analiza las características del sistema de evaluación federal anterior y posterior a esta innovación institucional, reconociendo los avances realizados, pero también los importantes desafíos aún pendientes.

Finalmente, presenta la experiencia iniciada por Evalúa DF a partir de 2008, con propuestas y procesos de planeación que implican diferencias de valores y de criterios de organización y decisión, constituyendo una práctica emergente novedosa frente al modelo tradicional federal.

## CAPÍTULO VII

### Metaevaluación

Este capítulo pretende revisar: 1) las razones que limitan el desarrollo de la evaluación; 2) los avances teóricos que se han producido en materia de meta-evaluación, entendida como la “evaluación de las evaluaciones”, con objeto de comprender y mejorar la calidad del trabajo de evaluación realizado; 3) algunas de las limitadas experiencias de metaevaluación actualmente disponibles fuera de México; y 4) describir los trabajos que se han realizado en los últimos cinco años en el área de desarrollo social del contexto mexicano, tanto a nivel federal como del Gobierno del Distrito Federal.

#### **Dificultades para establecer procesos de evaluación**

Si bien en los países escandinavos, y otros como Estados Unidos o Canadá, la idiosincracia, el grado de desarrollo y/o el sistema político, han permitido mayores o más rápidos avances, como hemos podido ver también existen naciones que, aun gozando de altos grados de desarrollo como Alemania, Reino Unido, Canadá o Francia, han enfrentado importantes resistencias para su aplicación.

Generalmente, las dificultades aumentan en países menos desarrollados como los latinoamericanos y se agudizan en casos como el mexicano, en que su historia política ha estado signada por un agudo presidencialismo y, hasta hace poco tiempo, por la existencia de un partido dominante, mismos que estuvieron lejos de desarrollar la rendición de cuentas a la ciudadanía y la cultura de la evaluación.

Como ya mencionamos, Nioche (1982) identificaba, en Francia, tres tipos fundamentales de problemas ligados al atraso relativo de una cultura de la evaluación en su país, que aparecen como muy pertinentes para el contexto mexicano: sociopolíticos, administrativos y metodológicos.

Los primeros son considerados los más importantes y, al mismo tiempo, los más difíciles de cambiar porque están enraizados en las culturas y tradiciones de los pueblos y sus gobiernos. La evaluación se vuelve amenazante en este contexto porque introduce la duda y la posibilidad de cuestionamiento entre los altos funcionarios, acostumbrados a comportamientos informales. Así, la demanda de evaluación no ha sido importante hasta fechas muy recientes, pero tampoco la oferta de trabajos. Otros factores que contribuyen a la explicación son: el débil rol del Poder Legislativo como evaluador de la acción gubernamental, el fuerte control de los medios de comunicación masivos por el Ejecutivo, el cuasimonopolio de la información económica y social por la administración pública y la ausencia de factores facilitadores, los que, por el contrario, han estado presentes en la experiencia de Estados Unidos, como su federalismo y tradición liberal (no intervencionista).

Los obstáculos administrativos incluyen: 1) la debilidad de la gestión por programas, que resulta el más evidente; 2) la organización del trabajo: el diseño de planes, programas y presupuestos con objetivos muy vagos y ausencia de metas cuantificadas que faciliten su evaluación; y una organización que no define con claridad una entidad administrativa encargada de la realización de una evaluación integral de los programas públicos; 3) los frenos de naturaleza financiera: la falta de recursos destinados a esta actividad, poco legitimada por su discutida aportación al bienestar social; y, 4) el rechazo de algunos funcionarios a una evaluación que consideran inútil o excesivamente restrictiva, dados los controles ya existentes.

Los problemas metodológicos tampoco pueden soslayarse. Como ya hemos visto, todas las técnicas empleadas en evaluación enfrentan serios cuestionamientos que disminuyen su credibilidad.

Para Nioche, las evaluaciones comienzan a desarrollarse en ámbitos caracterizados por un fuerte grado de incertidumbre o un elevado potencial de conflicto. Por otra parte, la evaluación sólo tiene sentido en una sociedad abierta a la libre expresión de los ciudadanos sobre la acción gubernamental, permitiendo una pluralidad de análisis, limitada por dos problemas relevantes: el acceso a la información y la desigualdad de los actores sociales en materia de recursos para realizar sus análisis.

Sin embargo, el análisis de políticas puede evitar dos grandes errores, generalmente presentes en la gestión pública, si se desarrolla en un contexto estructural y político que asegure la libertad de información y el pluralismo de los trabajos: considerar que todo es político y no utilizar las herramientas disponibles para comprender y mejorar la acción pública; o bien, querer utilizar herramientas técnicas para resolver problemas que son realmente de índole política.

Cuatro años después de realizado el artículo anterior, Deleau y otros (1986) percibían ya una atención creciente al tema y ofrecían cinco propuestas para lograr el desarrollo de la evaluación de políticas públicas en Francia: transformar a las políticas públicas en algo realmente evaluable, organizar las competencias, reforzar la evaluación de las políticas públicas en el seno de la administración, aumentar la capacidad de evaluación externa de la administración, facilitar su difusión en la sociedad, e identificar los temas prioritarios.

Para Monnier (1995), las condiciones que facilitan el desarrollo de la evaluación son:

- una relativa independencia de los organismos de evaluación, que permita una crítica responsable y un comportamiento profesional ético,
- una capacidad multidisciplinaria de los evaluadores,
- un acceso suficiente a las fuentes de información administrativa, y
- una pluralidad de órganos de evaluación que permita la emulación entre expertos y su empleo por los decisores como fundamento de sus negociaciones.

Muchas veces el resultado de una evaluación se “archiva” sin ser tomado en cuenta, a pesar de todos los cuidados que se hayan puesto en su preparación. Las razones pueden ser muchas: poca motivación de quienes solicitaron la evaluación, afectación de intereses, lenguaje poco accesible empleado por el evaluador, informe entregado en fecha poco oportuna, etcétera. A menudo, la evaluación entrega complejos informes donde los solicitantes buscaban respuestas sencillas, lo que crea registros distintos que obstaculizan la capitalización de las evaluaciones. En particular, se resalta la faceta política y psicológica de la evaluación que, al emitir un juicio sobre las finalidades, la eficacia y la ética de una acción pública, puede “hacer potencialmente vulnerables a los responsables de las políticas y las organizaciones públicas que son quienes contratan la evaluación” (Bañón, 2003: XXV).

Al respecto existe un interesante estudio sobre las representaciones, discursos y experiencias de quienes contratan una evaluación, resaltando su ambivalencia: por un lado manifiestan la importancia de su realización y por otro se perciben temores y aprehensión durante su aplicación (Martin y Ruffiot, 2001). La retórica del discurso y la institucionalización formal de la evaluación no garantizan el cambio de actitudes y conductas. En el mismo sentido se señala que: “La primera exigencia para que la evaluación sea posible y eficaz es que sea querida... Hay una evidente contradicción entre un sentimiento generalizado que define a la evaluación como indispensable y las reticencias a su concretización” (Duran, 1999: 188, traducción libre). La explicación del empantanamiento de la evaluación está tanto en “las expectativas que genera como los temores que despierta, y al mismo tiempo, las dificultades que implica su operacionalización” (Duran y Monnier, 1992, traducción libre). Un análisis psicológico detallado (exposición de los evaluados a diversas películas y análisis de sus reacciones) habría dado resultados tan negativos en cuanto a lo que la evaluación simboliza, también en el contexto francés, que nunca fue publicado.

Duran afirma que el conocimiento generado por la evaluación debe ser pertinente para quienes lo necesitan y susceptible de ser asimilado por ellos. El lenguaje habitual de los administradores es el de la gestión, no el de sus consecuencias lejanas. Además la evaluación cuestiona su acción y pone en peligro la existencia misma de la organización o programa evaluado. “Las ideas no deben su eficacia sólo a su fuerza lógica, necesitan relacionarse con el interés práctico de los actores y no deben ser demasiado extrañas a su experiencia” (Duran, 2002: IV, traducción libre). Finalmente, no hay que olvidar que la evaluación puede servir para redistribuir las relaciones de poder en la administración pública. Mientras no interese a sus usuarios potenciales, no tendrá impacto; evaluación y gestión necesitan relacionarse. Pero para lograrlo, es necesario garantizar la calidad de la evaluación, lo que nos relaciona con el tema de la metaevaluación.

## Conceptos e instrumentos de metaevaluación

Definimos provisoriamente a la metaevaluación<sup>1</sup> como un tipo particular de evaluación que enjuicia la calidad y el mérito de lo que hacen los evaluadores profesionales, persigue el perfeccionamiento de sus trabajos y promueve una mayor comprensión de los mismos.

A diferencia de las evaluaciones cuyo objeto de trabajo está constituido por las políticas o los programas aplicados, en el caso de la metaevaluación se toma como objeto de estudio a las evaluaciones realizadas, particularmente sus informes finales. En consecuencia, la metaevaluación se caracteriza por una mayor complejidad, determinada por el número y la naturaleza de las variables a considerar, y también profundidad, definida por sus propósitos y su objeto de trabajo (Díaz, 2002).

No ha sido frecuente que los principales autores que exponen la teoría de la evaluación la mencionen, tampoco se dispone de modelos desarrollados *ex profeso* ni de una epistemología y metodología para su construcción (Díaz y Rosales, 2003). Sus mayores desarrollos se localizan en el campo educativo y las primeras referencias que todos los estudiosos de la metaevaluación reconocen son los trabajos realizados por Michael Scriven (1967, 1969 y 1976), a quien se atribuye también la paternidad del nombre. En una presentación posterior del tema, Scriven se refirió a la misma en los términos siguientes: “la evaluación de las evaluaciones –indirectamente la evaluación de los evaluadores– representa una obligación ética y científica, cuando se trata del bienestar de los demás” (Scriven, 1991:228).

El autor subraya así que esta fase, usualmente terminal, del trabajo de evaluación responde a la responsabilidad exigible a todas las actividades humanas, misma que adquiere especial importancia en el caso de la evaluación, por lo que la metaevaluación representa el imperativo profesional de la disciplina, imprescindible para controlar el sesgo del evaluador (Scriven, 1976).

Stufflebeam (1974), define ocho premisas para diferenciar evaluación y metaevaluación:

<sup>1</sup> Es posible encontrar otro empleo de la palabra metaevaluación; en el caso de Cunill y Ospina (2003:444) es empleada para referirse a la evaluación a nivel “macro”, de las políticas públicas, por oposición de las efectuadas en los niveles “meso” referido a organizaciones y programas públicos, y del “micro” de desempeño de los funcionarios. También se encuentra este sentido en el trabajo de Sanín (1999) en la Comisión Económica para América Latina.

- La evaluación es la valoración del mérito, de ese modo, la metaevaluación significa la valoración o estimación del mérito del empeño evaluativo.
- La evaluación sirve a la toma de decisión y a la rendición de cuenta, por tanto, la metaevaluación deberá proporcionar información proactiva de soporte (o respaldo) a las decisiones que se toman en la conducción del trabajo evaluativo y deberá proporcionar información retroactiva para ayudar a los evaluadores responsables de su trabajo evaluativo anterior. Es decir, la metaevaluación debe ser formativa y sumativa.
- Las evaluaciones deben valorar la finalidad, diseño, implementación y resultados, así mismo la metaevaluación debe evaluar la importancia de los objetivos de la evaluación, la idoneidad de su diseño, la adecuación de la implementación al diseño de la evaluación y la calidad e importancia de los resultados de la evaluación.
- La evaluación debe proporcionar una información descriptiva, crítica y además recomendaciones apropiadas. Igualmente, la metaevaluación debe describir y juzgar el trabajo evaluativo, también debe recomendar cómo las evaluaciones pueden ser perfeccionadas y cómo los hallazgos pueden ser usados apropiadamente.
- La evaluación debe servir a todas las personas, tanto aquellas que están involucradas en ella como a quienes han sido afectadas por el programa, consecuentemente la metaevaluación debe servir a los evaluadores y a todas las personas que están interesadas en el trabajo de éstos.
- La evaluación debe ser conducida por personas internas (evaluación formativa) y externas al programa (evaluación sumativa). Por tanto, se debe conducir una metaevaluación formativa y obtener juicios externos del valor global de las actividades evaluativas realizadas.
- La evaluación involucra el proceso de esbozar los asuntos a los que se deben dirigir, el proceso de obtención de la información necesaria y el uso de la información para la toma de decisión y la rendición de cuentas. Los metaevaluadores deben realizar tres pasos: determinar los asuntos específicos a que debe dirigirse la metaevaluación; acopiar, organizar y analizar la información necesaria y, finalmente, emplear la información obtenida en la toma de decisión apropiada y en la rendición de cuenta.
- La evaluación debe ser técnicamente adecuada, útil y efectiva. La metaevaluación debe satisfacer los mismos criterios (Stufflebeam, 1974, citado en Dopico, s/f: 9-10).

House (1980, citado en Díaz, 2002) dedica al tema una sección de su libro titulado *Evaluación, ética y poder*. Fundamenta el trabajo del metaevaluador en la percepción selectiva “de las disonancias que se presentan en la evaluación y en su informe”, misma que podría analizarse a partir de dos enfoques: las opciones “duras” basadas en la cuantificación de variables, relaciones de causalidad, criterios de eficacia y eficiencia, etcétera; y las opciones “suaves” con gran cantidad de variables cualitativas interactivas implicadas en un trabajo hermenéutico de descripción, comprensión e interpretación de textos. Las primeras opciones podemos vincularlas a definiciones de metaevaluación de contenido práctico y operativo; mientras que las segundas se asocian con la búsqueda más profunda de sentido y explicación de los procesos evaluativos.

House se inclina tanto a la utilización de ejercicios sistemáticos de hermenéutica aplicados al texto (búsqueda de relaciones de identidad, contradicción, reciprocidad, transitividad, etcétera) como al empleo de técnicas cualitativas de manejo de datos (análisis, síntesis, comparación, causalidad, etcétera).

Schwandt y Halpern (1988) proponen dos modelos de metaevaluación. El primero se trata de la revisión por el metaevaluador, sólo o conjuntamente con el evaluador del programa, de las etapas desarrolladas en el proceso, entre las que resaltan las de preparación del trabajo: análisis de su factibilidad, negociación del contrato, planeación de actividades, etcétera. El segundo, se refiere al proceso mismo de investigación y consta de una lista de comprobación para garantizar las condiciones de confiabilidad y validez de los resultados.

La posición de estos autores nos parece de particular interés por dos asuntos que plantean, a diferencias de los demás, y que se relacionan con nuestra propuesta: 1) la posibilidad de que el metaevaluador no realice su trabajo solo, sino que pueda interactuar con los evaluadores, y 2) la necesidad de analizar, no únicamente el proceso y los métodos de evaluación aplicados, sino también algunos aspectos previos y posteriores, vinculados al funcionamiento del sistema o contexto en que el trabajo se inscribe.

Scriven (1991), propone la aplicación de su *Key Evaluation Checklist*, compuesta de las siguientes 13 dimensiones o aspectos a revisar, como instrumento de evaluación que también puede ser usado en la metaevaluación:

1. Descripción: diseño, equipo evaluador, calendario del estudio, fuentes de información, entre otras.

2. Antecedentes y contexto: responsable de la solicitud de evaluación, razones, evaluaciones previas, etcétera.
3. Consumidor: clientes y destinatarios efectivos y potenciales.
4. Recursos: insumos que pueden limitar el diseño y ejecución de la evaluación.
5. Valores: criterios de mérito en función de las necesidades del cliente y las normas profesionales.
6. Proceso: desarrollo de la evaluación y problemas encontrados.
7. Resultados: informes y sus efectos previsibles por el evaluador así como impactos del proceso de evaluación.
8. Costes: monetarios y no monetarios, así como sus costes alternativos.
9. Comparaciones: utilización de juicios validados, el azar, etcétera.
10. Posibilidades de generalización: posibles usos en diseños futuros, otros programas, otros lugares, entre otros.
11. Relevancia: síntesis de todo lo anterior que permite valorar el mérito de la evaluación realizada.
12. Recomendaciones: se pueden o no hacer en torno a la continuidad del programa y sus posibles ajustes.
13. Informe del evaluador.

Stufflebeam y Shinkfield (1993) se acercan al concepto de metaevaluación definido por Schwandt y Halpern (1988) para incluir la valoración “de todo lo que representa y hace la profesión”, con objeto de garantizar que la evaluación cumpla con estándares de adecuación técnica, costo-efectividad y utilidad, incluyendo la planeación de la evaluación, la investigación evaluativa, el desarrollo organizativo, los servicios de evaluación y la utilización de sus resultados. Retoman la lista de Scriven y otra similar preparada por Stufflebeam (1974) en la Western Michigan University y elaboran una versión depurada que diferencia entre el cliente que encarga la evaluación y sus consumidores, y que agrega nuevos elementos como la función que cumple el evaluado, el sistema que utiliza para distribuir sus servicios, el significado o síntesis de todo lo analizado y la metaevaluación que puede incorporarse al propio informe de evaluación.

La metaevaluación también ha sido reconocida por especialistas más recientes como Monnier (1995), Patton (1997), Weiss (1998), Bañón (2003), entre otros autores, y por organizaciones como el Joint Committee on Standards for Educational Evaluation (1994). El segundo de los mencionados plantea que la

metaevaluación debe analizar las cuatro categorías básicas que componen a toda evaluación: datos, interpretaciones, valoraciones y recomendaciones.

Díaz cita algunos trabajos adicionales<sup>2</sup> entre quienes resalta Santos Guerra (1999), quien además de elaborar un inventario sistemático de los elementos a comprobar (necesarios pero no suficientes para evaluar el trabajo realizado), elaboró preguntas relevantes para los grandes momentos de la evaluación:

- constitución: origen de la evaluación, negociación de sus términos, etcétera,
- justificación: rigor metodológico, validez, tratamiento de la información, entre otros,
- presentación: principalmente, estructura lógica del informe, extensión, coherencia y claridad,
- negociación: rapidez, intermediación, jerarquización en la entrega del informe;
- valoración: motivaciones, intenciones, presiones, actitudes de los participantes, y
- difusión: distribución de la información y resultados, demoras y aduleteraciones.

Se comparte con Díaz la idea de que la metaevaluación puede enfocarse en forma restringida a los datos y hechos, que constituyen el contenido técnico de la evaluación (procesos y resultados) y que la mayoría de los autores señalan como su contenido, y operacionalizarse mediante listas de comprobación. Pero también puede trascenderlos para tomar en cuenta todos los factores que la afectan. Para Díaz dichos factores incluyen principalmente: evaluadores,

<sup>2</sup> Entre ellos incluye cuatro artículos relevantes y una bibliografía sobre el tema. Los primeros son: Cook y Gruder (febrero de 1978), "Meta-evaluation Research", *Evaluation Quarterly*, 2(1); Reineke y Wech (1986), "Client-centered meta-evaluation", *Evaluation Practice*; Santos Guerra (1999), "Metaevaluación: rigor, mejora, ética y aprendizaje", en Jiménez Jiménez, Bonifacio, "Evaluación de programas, centros y profesores", Síntesis, Madrid; y J.F. Angulo (1993), "Metaevaluación del programa de educación de diabéticos", en Soriguer *et al.*, citado por Bonifacio Jiménez, *Evaluación de programas, centros y profesores*, Síntesis, Madrid. La bibliografía corresponde a Patricia Rogers (comp.), *Metaevaluation bibliography*, Program for Public Sector Evaluation, Royal Melbourne Institute of Technology, 1997.

modelos, categorías, fundamentos metodológicos y epistemológicos, paradigmas científicos, ideología, filosofía, usos de lenguaje, codificación, conducta del evaluador, valores en juego.

Sin embargo, el autor reconoce que generalmente no se cuenta con información de este cúmulo de aspectos, ya que la metaevaluación se realiza con base en los informes oficiales del evaluador, por lo que recurre a proponer el método hermenéutico para desentrañar la ideología, la concepción del mundo y la epistemología que fundamentan los modelos y categorías orientadores del proceso evaluativo. Aclara que dicho método “consiste en repetir sistemática y cíclicamente, a lo largo de todo el proceso, el ejercicio representado por las cuatro funciones: análisis estructural, interpretación, comentario y crítica” (Díaz, 2002:329).

Santos Guerra refiere la necesidad de interrogarse sobre la naturaleza de la iniciativa, la finalidad de los procesos, la utilización de sus resultados, el acceso democrático a los informes, o sea, el conjunto de factores hacia el que parecían apuntar las anteriores propuestas de Stufflebeam y Shinkfield y de Schwandt y Halpern.

Para García Sánchez (2003 y 2004), trascender las características más evidentes de las evaluaciones implica desentrañar sus causas mediante el estudio de las decisiones de los actores y sus motivaciones, recuperando las contribuciones pluralistas, lo que coadyuva a la comprensión de la forma en que se distribuye el poder en estos procesos. Propone dar seguimiento a las decisiones de mayor relevancia en el proceso de evaluación:

- selección del equipo evaluador y contrato de evaluación,
- enfoque metodológico,
- selección de actores participantes,
- objeto y objetivos de la evaluación,
- criterios de valor,
- técnicas de recolección y análisis de datos,
- contenidos y estilo del informe de evaluación,
- difusión y/o publicación del informe, y
- utilización del proceso y los resultados de la evaluación.

En este trabajo consideramos que este aditamento se conforma por factores de carácter organizacional, que es posible reconstruir echando mano de la técnica

de entrevista aplicada a informantes-clave, como complemento del análisis de informes, una de las técnicas que vimos sugería Abate (2000) para recoger la percepción de los diferentes actores vinculados al proceso.

Desde la versión comprensiva de Stake (2006), la metaevaluación debe construirse a partir de la experiencia de los actores, en diálogo con las personas y sus opiniones sobre los hechos. A ello, Murciano (2007) agrega que lo anterior debe hacerse aprehendiendo el contexto en el que actúan los equipos evaluadores para entender sus decisiones e interpretaciones.

Recientemente ha aparecido un nuevo documento sobre metaevaluación, desarrollado en la Universidad de La Habana, que también retoma los aspectos organizacionales. Coincide en que la metaevaluación persigue el perfeccionamiento de la calidad de la evaluación, mide la veracidad, viabilidad, objetividad y fiabilidad de la información recopilada y generada, lo que implica evaluar el proceso, el sistema de evaluación en cuestión y su impacto real, entendido como la influencia perceptible que la evaluación ejerce sobre las actividades o actitudes de individuos o grupos (Dopico, s/f).

En particular, resulta relevante investigar las debilidades o deficiencias del trabajo de evaluación realizado, para lo cual Dopico nos ofrece un amplio conjunto de aspectos en los cuales podemos detectarlos, muy similar a las listas de chequeo presentadas en Cardozo (2006b):

- La definición de los objetivos que se persiguen con la evaluación
- La elección del momento en que se realiza la evaluación
- La duración o premura con que se realice la evaluación
- La predeterminación de los parámetros a evaluar
- El sistema de puntaje que se utilice
- La pertinencia del sistema de evaluación diseñado para que sea aplicado a todos los programas
- La identificación del sistema de evaluación diseñado con lo que se entiende como patrón de calidad de los programas
- El cumplimiento por los evaluadores seleccionados con aquellas aptitudes y actitudes establecidas para ejercer como evaluadores: justeza, objetividad, honestidad, capacidad, etcétera.
- La organización del trabajo de los evaluadores
- El nivel de participación de los interesados en el proceso de evaluación
- El clima preponderante durante la evaluación

- El apoyo institucional para que el proceso de evaluación se lleve a cabo en las condiciones requeridas
- Las técnicas empleadas por los evaluadores para la recolección de la información
- La calidad de los documentos del programa presentados por los evaluadores
- La suficiencia de medios y recursos con que se cuenta para la realización de la evaluación
- La utilización de dichos recursos
- La riqueza de la información aportada por los encuestados y entrevistados
- La calidad del informe final presentado por los evaluadores (Dopico, s/f:11).

Finalmente, Dopico agrega una nueva utilidad práctica de la metaevaluación: dar respuesta a las discrepancias surgidas durante la evaluación, cuando la parte evaluada considera que el proceso no fue justo y transparente y tiene quejas sobre cualquiera de sus aspectos (los métodos y técnicas empleados, la ética y profesionalidad del evaluador, etcétera).

Un segundo nuevo trabajo sobre metaevaluación, nuevamente en el ámbito educativo (Marabotto, 2005), reconoce que la evaluación está impregnada de valores, ideología, intereses y juegos de poder, de la ambigüedad de sus procesos y resultados, por lo que depende de la rigurosidad de sus evidencias, pero también de su legitimidad (consistencia, credibilidad, transparencia y aceptación) y recomienda documentar sistemáticamente las prácticas.

El trabajo llama la atención en el sentido de que el proceso de evaluación puede tornarse ingenuo, viciado, irrelevante o burocrático por lo que se justifica la revisión epistemológica, ética y metodológica de la evaluación. Aporta también la consideración de un aspecto poco abordado: el de los procesos emocionales que operan durante la evaluación, por lo que se recomiendan verificaciones basadas, por ejemplo, en una doble mirada o la organización de grupos cuasiterapéuticos centrados en el diálogo, el consenso y el acompañamiento (Cardozo, 2009).

Tratando de esquematizar la revisión de conceptos e instrumentos realizada, identificamos:

- metaevaluaciones dirigidas exclusivamente al análisis del informe final; otras orientadas a la revisión de los procesos técnico-metodológicos implicados en la investigación; o, aquellas que se enfocan a los anteriores en el marco

de un contexto político y una determinada estructura y funcionamiento organizacional,

- metaevaluaciones con enfoque racional, cuantitativo y de perspectiva positivista basado en el análisis estadístico de datos; con enfoque fenomenológico, holista, poco estructurado y cualitativo, fundamentado en la hermenéutica y el estudio de los aspectos políticos, organizativos o éticos; o, con enfoque mixto, y
- metaevaluaciones que utilizan diversas listas de indicadores o aspectos a revisar, propuestas por distintos expositores del tema.

Por otra parte, Bustelo (2001 y 2002) plantea que existen autores y organizaciones que confunden el concepto de metaevaluación con otros como la síntesis evaluativa y el meta-análisis. Se trata del caso de la Evaluation Research Society que la considera uno de los seis tipos de evaluación que reconoce y la define como una forma de volver a analizar los resultados de una o varias evaluaciones para determinar lo que se ha aprendido en torno a esa política o programa. Vedung (1997) incluye en el concepto de metaevaluación la síntesis de resultados de diferentes estudios de evaluación y Ballart (1992) expresa: “consiste en el análisis del resultado de varias evaluaciones, y por otro lado, en la revisión de la más amplia investigación desarrollada en un cierto periodo”.

Para la autora los tres conceptos que se mezclan son diferentes:

- la síntesis evaluativa resume los resultados de múltiples evaluaciones efectuadas a políticas o programas similares, y sólo se interesa en la metodología como la base que permitió llegar a los resultados (no analiza los procesos),
- el meta-análisis consiste en un procedimiento estadístico para comparar los hallazgos de diferentes estudios que confluyen en un mismo objeto, identificar tendencias y realizar generalizaciones, que ha sido aplicado a las evaluaciones (Glass, 1986), y
- la metaevaluación no se enfoca a los resultados de las evaluaciones sino a los procesos implicados en la realización de una o varias de ellas y, cuando incluye a los primeros, no es para sintetizarlos sino para revisar si cuentan con evidencia suficiente, si la metodología ha sido adecuada, si las interpretaciones, juicios y recomendaciones fueron lógicamente diseñados, etcétera. Cumple dos funciones básicas: la de descripción, análisis y valoración de los procesos de evaluación, y la de seguimiento, fomento y control de la calidad de las evaluaciones, mismos

que también permiten asegurar la credibilidad de la evaluación en contextos muy politizados.

Bustelo agrega que la literatura muestra una tendencia a metaevaluar una única evaluación,<sup>3</sup> ya que son raros los procesos de evaluación que se analizan y comparan simultáneamente debido a que el trabajo se guía más por la lógica y tiempos de la acción pública que por los de la construcción del conocimiento.

En conclusión, dentro del presente libro se comparte, pero se amplía la definición presentada al inicio ya que se considera que la comprensión y el perfeccionamiento de la evaluación debe llevarnos a enjuiciar el valor y el mérito no sólo de los trabajos de los evaluadores profesionales, sino el de todo el sistema en que se organiza y desarrolla el anterior, incluyendo el papel que desempeñan otros actores o instancias que inciden en la calidad del proceso desarrollado y del informe final logrado.

Esto implica no reducir la metaevaluación al estudio de los métodos y procesos aplicados por el evaluador, sino recuperar fundamentalmente las propuestas que desde la teoría de la organización y el enfoque pluralista sugieren tomar en cuenta el contexto inmediato en que la evaluación se desarrolla, considerada ésta como un proceso social complejo. Por otra parte, significa que ninguna disciplina ni enfoque puede pretender ser el único adecuado para realizar una metaevaluación, por lo que su análisis debe ser de tipo multidisciplinario y mixto.

### **Criterios de metaevaluación**

Aún son muy escasas las propuestas de criterios específicos para la metaevaluación. House (1980, citado en Díaz, 2002) propuso criterios muy generales, poco operativos, que recuperan en forma amplia los procesos relacionados con la evaluación, tales como veracidad, autonomía, credibilidad, participación, equidad, poder, objetividad, etcétera, los cuales considera que deben coadyuvar al logro de una mayor justicia social.

Patton (1997) señala que el debate entre cuantitativistas y cualitativistas sirvió para mostrar que no hay criterios universalmente válidos en todos los casos; sin

<sup>3</sup> Sin embargo, no es lo que encontramos en nuestra investigación.

embargo, también plantea que existen algunos de carácter general que pueden utilizarse como referencia: conveniencia, utilidad, factibilidad, credibilidad y relevancia.

Scriven (citado en García, 2009) elabora, a partir de diferentes listas, una síntesis de aspectos específicos denominada Meta-Evaluation Checklist, que incluye las características deseables a verificar en una metaevaluación: estudio comprensivo y comprensible, creíble, ético, factible, suficiente, oportuno, seguro, válido, sensible desde el punto de vista político, que explicita los estándares de valor utilizados y su justificación, y costo-eficaz.

En cambio, Stufflebeam (1974, citado en Centeno, 2002) recupera los principales criterios metodológicos y epistemológicos, en relación exclusiva con los procesos de investigación: validez interna, validez externa, confiabilidad, objetividad, relevancia, importancia, alcance, credibilidad, tiempo, difusión, y relación costo-efecto.

Schwandt y Halpern (1988) también identifican criterios específicos de calidad de una metaevaluación, sin tomar en cuenta su contexto: el rigor de la metodología utilizada y su capacidad de convencimiento basada en la bondad de los resultados hallados en la evaluación.

La primera se refiere a los procedimientos utilizados para conocer el objeto de estudio (evaluación): coherencia epistemológica, metodológica y tecnológica, adecuación a su contexto, relevancia para la política o programa evaluado, definición de criterios preestablecidos de evaluación, técnicas de recopilación, análisis de datos, participación de actores relevantes y relación costo-beneficio de la evaluación.

La segunda trata de conseguir la coherencia lógica, el señalamiento independiente y la vinculación de los cuatro niveles definidos por Patton (1997), en que cada nivel se basa en el anterior: datos, interpretaciones, valoraciones y recomendaciones.

Schwandt y Halpern definen los siguientes criterios para valorar los datos: relevancia, confiabilidad, suficiencia, representatividad y temporalidad. Las interpretaciones y explicaciones obtenidas a partir de los datos que cumplen los criterios anteriores serán adecuadas si su análisis es correcto y suficiente. Las valoraciones deben hacerse en función de los criterios preestablecidos en acuerdo con los diferentes agentes de la evaluación. Las recomendaciones deben basarse en los niveles previos y considerar el contexto en que se desarrolló la

acción evaluada, ser realistas, claras y se recomienda que se acompañen de una estimación de sus costos para facilitar su posible implementación.

Finalmente, las normas más utilizadas como referencia para la metaevaluación son las del Joint Committee on Standards for Educational Evaluation, que aparecen ordenadas en cuatro grupos: utilidad, factibilidad, corrección/legitimidad, y precisión (Anexo 1). Las mismas son reconocidas por su especificidad, claridad, orden y estructuración, pero cada metaevaluación puede definir sus propios criterios (Bustelo, 2001).

García (2009) menciona que lo importante es valorar la capacidad explicativa de la investigación evaluativa, por lo que los criterios se deben determinar en función de los objetivos de la investigación y de la perspectiva del analista.

En conclusión, existe una amplia gama de criterios a seleccionar en cada caso, mismos que incluyen:

- los de carácter general, relativos a los procesos de investigación: objetividad, validez, confiabilidad, representatividad, suficiencia, veracidad, relevancia, relación costo/beneficio, etcétera.
- los específicos de la metaevaluación: autonomía, credibilidad, utilidad, factibilidad, oportunidad, participación, legitimidad, etcétera.

Los criterios se combinan en índices para sintetizar una medida de la calidad de los trabajos, tal como sugiere Abate (2000).

### **Tipos de metaevaluación**

A las metaevaluaciones pueden aplicarse todas las clasificaciones de evaluaciones, ya revisadas.

Los enfoques *cuantitativistas* enfatizan el análisis de los aspectos metodológicos y, en casos extremos, identifican como modelo de calidad del trabajo el empleo de diseños experimentales (Campbell y Stanley, 2005).

Las corrientes *cualitativistas* enfatizan los aspectos éticos y políticos de la investigación, a veces incluso con descuido de las cuestiones metodológicas (House, 1980; Scriven, 1991).

Las *metaevaluaciones de diseño* deben analizar el contexto evaluativo, la claridad de sus objetivos, la inclusión de las variables relevantes, la justificación del esfuerzo, su costo, el tiempo disponible, su contribución potencial a la mejoría del programa, la participación de los actores involucrados, la disponibilidad de la información necesaria, la previsión de técnicas adecuadas de recolección y análisis de datos, de presentación y difusión de informes, etcétera.

Las *metaevaluaciones de proceso* analizan la puesta en práctica de la evaluación antes diseñada: ajuste al calendario, disponibilidad oportuna y uso eficiente de recursos, calidad del equipo evaluador, aplicación de los métodos y técnicas previstas, problemas encontrados, entre otros.

Algunos autores consideran como resultados de la *metaevaluación* tantos sus *productos* como sus *impactos*. En la primera, se revisa la calidad de los productos típicos de una evaluación, que incluyen datos, interpretaciones, valoraciones y recomendaciones, así como la forma en que son recuperados en los informes. También se analiza si éstos se entregaron a tiempo y se difundieron.

La segunda se refiere a los efectos que provoca la evaluación, su utilización efectiva: modificaciones al programa, contribuciones a la política en que se enmarca, procesos de aprendizaje, etcétera (Bustelo, 2001). Weiss (1998) define cuatro posibles impactos o usos de la evaluación: instrumental (para su uso en la toma de decisiones), conceptual (para cambiar la comprensión de lo que el plan o programa hace), persuasivo (cuando moviliza apoyos para efectuar los cambios requeridos por el programa), y de influencia en redes de políticas, programas o evaluadores, ajenos al programa evaluado.

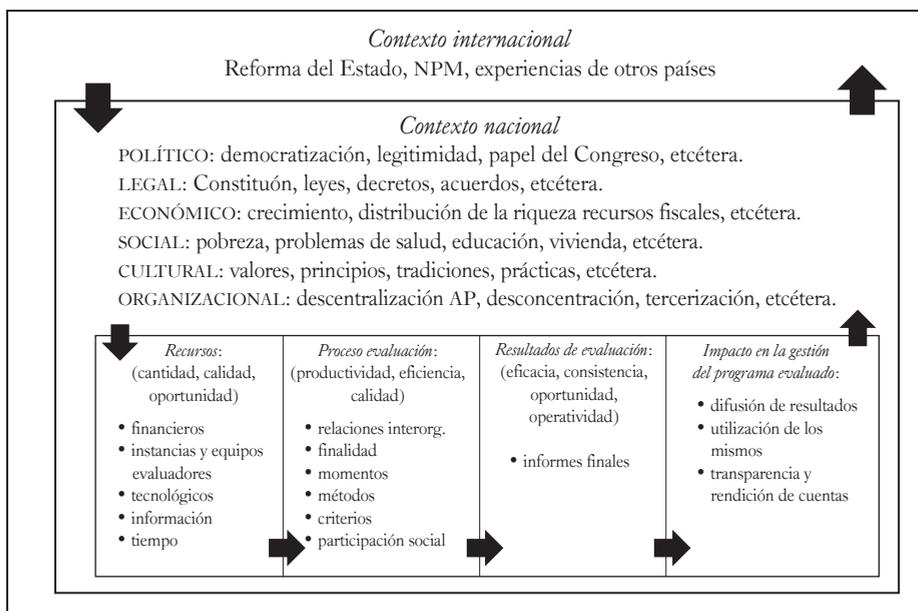
Por su parte, Lascoumes (1998) los clasifica en impactos positivos, o sea los que generan una dinámica de cambio, e impactos negativos (que no generan cambios o incluso provocan resistencias).

Entre las consecuencias positivas identifica: consecuencias técnicas o pragmáticas (nuevos conocimientos, integración de los dispersos, dinámicas de capacitación); acciones de reencuadre de la acción pública (clarificación de competencia entre actores, mayor coherencia en la toma de decisiones, adaptación de recursos); consecuencias heurísticas sobre los actores individuales o institucionales implicados, relacionadas con su implicación en el dispositivo de evaluación, por lo tanto, limitada a un pequeño círculo, poco difundible, temporal; y, consecuencias políticas (reformulación del problema, reingreso a la agenda, cambios estratégicos).

Por su parte, las consecuencias negativas pueden incluir: falta de apropiación de los resultados de la evaluación por las instancias políticas y/o administrativas; rechazo de los resultados por su solicitante debido a cambios en el contexto político (más rara), mala interpretación de la solicitud, o conclusiones imposibles de digerir para la élite gobernante; desnaturalización de los datos y las recomendaciones de la evaluación; e, imposibilidad de un actor político que forma parte del proceso de evaluación de introducir cambios en función de los resultados obtenidos debido a la falta de recursos materiales o simbólicos.

En síntesis, una metaevaluación integral, al igual que se planteó en el caso de las evaluaciones en general, debe cubrir tanto el diseño, como el proceso, productos e impactos, y es conveniente que combine enfoques cuantitativos y cualitativos con perspectiva interdisciplinaria (Diagrama 4).

DIAGRAMA 4  
*Metaevaluación integral*



Fuente: elaboración propia.

## Experiencias de metaevaluación fuera de México

Bustelo (2001 y 2004) plantea que, a pesar del desarrollo alcanzado por la evaluación en Europa, aún hay muy poca investigación que sistematice y analice estas experiencias (metaevaluación). En los países en que la cultura y la práctica de la evaluación están más desarrolladas se ha insistido, en los últimos años, en la necesidad de realizar investigación que aproveche los estudios ya realizados. Algunos antecedentes pueden encontrarse en Reino Unido, Canadá o Bélgica.

En el presente libro se mencionan los trabajos a los que se tuvo acceso: varios realizados en Francia y España, además del canadiense ya mencionado, y los identificados en América Latina.

Casi todos los trabajos extranjeros refieren sus técnicas de recolección de la información, pero ninguno explicita el tratamiento posterior que dio a la misma para arribar a sus conclusiones.

Constituye un primer trabajo de metaevaluación el realizado por Lacasse (1993) en Canadá, referido dentro del segundo apartado de este capítulo. Para su estudio, inició un diagnóstico de la situación, basado en la revisión de 37 informes de evaluación del periodo 1988-1990, complementado con entrevistas confidenciales aplicadas a funcionarios, consultores y clientes de las evaluaciones, con base en una muestra intencional en que seleccionó sólo los trabajos de mayor calidad y dirigidos por personas de reconocida competencia.

Lascoumes (1998) relata una metaevaluación francesa que dio origen a la realización del coloquio "Evaluation pluraliste des politiques publiques: enjeux, pratiques et produits". Su objetivo era enfocarse al conjunto de trabajos colectivos que permitían la materialización de la evaluación: diversidad de prácticas, herramientas, modelos teóricos y metodologías. Especialmente se trató de saber si los evaluadores habían podido conciliar y cómo los objetivos de racionalización y coherencia de la política con los de su legitimación y aceptación, ya que el modelo francés de 1990 conservaba elementos de una evaluación técnica o gerencial, al mismo tiempo que introducía el enfoque pluralista, negaba prioridad al saber de los expertos y abría el proceso a los "profanos implicados".

Como objeto de observación se seleccionaron seis evaluaciones con diferentes encuadres para averiguar si éstos provocaban consecuencias en los procesos y productos de la evaluación: cuatro evaluaciones presentaban el modelo CIME-

CCE;<sup>4</sup> la quinta, aunque también era interministerial, se realizó en un encuadre estrictamente administrativo;<sup>5</sup> y la sexta no fue interministerial ni nacional, se trataba de la evaluación de una política pública local, realizada en un encuadre similar al CIME-CCE, aunque con algunas diferencias significativas.<sup>6</sup> Las cuatro primeras pudieron llegar a conclusiones prudentemente generalizables; las dos restantes hicieron un aporte distinto, sin pretensiones de representatividad de ningún tipo.

Analizaron todos los documentos posibles, producidos por las instancias de evaluación, así como aplicaron entrevistas a sus miembros y a algunos actores externos (solicitantes y encargados de la evaluación).

Dedujeron que la construcción metodológica de cada instancia dependía de cuatro factores: 1) el contexto político en que se seleccionaba al responsable de la evaluación y al objeto y problemática que ésta cubriría; 2) el tipo de relación que se establecía entre los diferentes grupos e intereses al interior de la instancia de evaluación 3) el respaldo científico brindado por el CCE o instancia equivalente, que debería apoyar en forma continua el proceso metodológico, pero sólo participa al inicio y al final del mismo; y 4) las instancias, que constituyen intelectuales colectivos que se diferencian por sus dinámicas internas y sus producciones.

Concluyeron que la evaluación no es un factor de rediseño de las políticas públicas; su impacto es más sensible y profundo como forma de esclarecer problemas y lograr un aprendizaje que podría transformarse en un proceso de cambio al interior de la administración. Otra de sus conclusiones fue que el modelo pluralista francés enfrentaba tres tipos de problemas: 1) la dificultad para superar la diferencias entre la lógica del solicitante, que busca controlar el proceso, con la de un pluralismo real; 2) la tendencia a ocultar los objetivos de la política a evaluar, que no siempre pueden ser descubiertos por la instancia de evaluación; y 3) las consecuencias efectivas de la evaluación, más indirectas que directas, y raramente sostenidas por quienes iniciaron el proceso.

<sup>4</sup> La informática en el Estado, La inserción de los adolescentes en dificultades, La rehabilitación del hábitat social, y El desarrollo de infraestructura en las zonas húmedas.

<sup>5</sup> La estrategia de formación de trabajadores sociales, realizado por el Ministerio de Asuntos Sociales.

<sup>6</sup> Política pública regional de turismo rural en la región de Bretaña.

Bustelo (2001), en su trabajo aplicado a la metaevaluación de políticas de igualdad de género en España, analiza once planes y programas en el tema. Utiliza técnicas de recopilación y análisis documental; entrevistas abiertas semiestructuradas aplicadas a los responsables de los organismos involucrados; y entrevistas estructuradas, personales y telefónicas, realizadas a los responsables técnicos de la evaluación en los mismos organismos. Señala que existe un claro déficit de valoraciones globales en las evaluaciones aplicadas y que ninguno de los informes revisados en forma detallada contiene un apartado de recomendaciones, lo que influye en la falta de utilidad y de impacto de dichos trabajos.

Lajouidie y Toulemonde (2002) analizaron once evaluaciones realizadas en la zona de Limousin, Francia, entre 1994 y 1998, en relación con sus procesos y la calidad de sus informes, y seleccionaron cuatro de ellas para profundizar en la utilidad de las mismas. Los temas revisados incluyeron: calidad de la solicitud de evaluación, conducción de la evaluación, calidad del informe y explotación de sus resultados. Para la revisión de los métodos de evaluación empleados utilizaron la aplicación de entrevistas directas, por teléfono y por correo electrónico. También revisaron los principales cambios acontecidos en el periodo. Entre sus conclusiones resalta la poca utilización de los resultados de las evaluaciones.

Larrú (2003) realizó una metaevaluación del sistema de evaluaciones de EuropeAid para valorar la calidad y la utilidad de los informes de evaluación externa contratados por su Unidad de Evaluaciones de la Unión Europea. Concluye que existe una gran variedad de tipos de evaluación aplicados (intermedias, *ex-post*, sumativas, formativas, etcétera), que las recomendaciones son pertinentes aunque a veces demasiado generales, algunos informes demasiado extensos (más de 100 páginas), se presentan riesgos de minimizar costos a expensas de la calidad, y dudas sobre la influencia real de las recomendaciones efectuadas y su impacto en la asignación de recursos.

Plantea además algunos dilemas entre los que incluye que las evaluaciones se dirigen a quienes las financian y elaboran unilateralmente los TR, por lo que las mismas ayudan muy poco a los beneficiarios de los programas y nunca enjuician la actuación de los donantes. También identifica futuras líneas de investigación que requerirían la aplicación de encuestas a funcionarios para complementar los datos incluidos en los documentos que analiza (costo de las evaluaciones, TR, etcétera).

También en España existe un trabajo de metaevaluación, de enfoque cualitativo, efectuado por García Sánchez (2003 y 2004). Fue aplicado a tres de las evaluaciones más destacadas de programas y proyectos de reforma educativa, realizadas en el periodo 1982-1992. Se basó en el enfoque de actores y trató de encontrar las causas que explican las decisiones adoptadas por los mismos.

Sus principales conclusiones incluyen: 1) los intereses que los actores no mencionan llegan a ser los que guían realmente los objetivos de las evaluaciones, 2) la mayoría de las decisiones fueron tomadas por un número reducido de actores, 3) las decisiones sobre el enfoque metodológico son las de mayor relevancia porque influyen en el resto de las decisiones críticas de la evaluación, 4) las evaluaciones analizadas no cuestionaron los programas y sus premisas, 5) tuvieron un escaso impacto en el perfeccionamiento y reorientación de los programas, 6) no se detectó correlación alguna entre el grado de pluralismo de las evaluaciones y la utilidad de sus resultados, etcétera.

En síntesis, cinco de las seis metaevaluaciones revisadas concluyeron que los resultados de las evaluaciones son muy pocos utilizados y el sexto (Larrú) manifiesta dudas al respecto. En el mismo sentido se expresan Díaz y Rosales (2003) con base en las evaluaciones que revisan en el contexto costarricense.

Como ha podido verse hasta el momento, es en el continente europeo donde se han identificado más casos de metaevaluación. En América Latina se identificó un primer trabajo de tesis, desarrollado por Centeno (2002) en la Universidad de Costa Rica, que trata de realizar una metaevaluación del Programa antidrogas y salud sexual y reproductiva para adolescentes, aplicado en Jamaica, en 1999-2002.

A pesar de que, según Bustelo (2001), existe una tendencia a metaevaluar una sola evaluación, los casos presentados hasta el momento se referían todos a múltiples evaluaciones; en cambio, el trabajo de Centeno, sólo metaevalúa un programa. Si bien el mismo plantea que analizará en forma integral todo el proceso (no queda claro si de evaluación o de gestión del programa) y que le interesan tanto los aspectos cuantitativos como los cualitativos, su trabajo se refiere únicamente en los aspectos técnicos, que son analizados en forma cuantitativa: las estrategias de análisis estadístico utilizadas durante la evaluación para la obtención de datos, su tratamiento e interpretación.

Realiza su trabajo debido a una discrepancia inicial con la forma en que se obtuvieron los resultados de la evaluación, rebasando el concepto de

metaevaluación que hemos manejado, ya que no se limita a analizar críticamente los procesos desarrollados, sino que propone y aplica nuevos métodos de tratamiento a los datos. En realidad, realiza una metaevaluación de los aspectos metodológicos y agrega una segunda evaluación del programa (ajena ya al concepto de metaevaluación), a partir de los mismos datos, aplicando nuevos métodos, lo que le permite obtener resultados también diferentes.

Entre las conclusiones de su metaevaluación resalta la ausencia de: una valoración inicial de las necesidades de los beneficiarios; una teoría del programa que relacione los insumos y actividades con el resultado esperado; pruebas de validez y confiabilidad del cuestionario aplicado; y utilización de un método de tratamiento ajustado a los datos recogidos.

### **Experiencias mexicanas de metaevaluación**

En el caso mexicano, se realizó un primer esfuerzo en 2006, recuperando la experiencia internacional reseñada, para analizar las evaluaciones aplicadas a los programas federales de desarrollo social en 2002-2003 (Cardozo, 2006). La misma pretendió cubrir los objetivos revisados (describir, analizar y valorar los procesos de evaluación y la calidad de sus resultados) y desarrollar los determinantes de carácter organizativo. Para ello analizó los informes finales de evaluación, los procesos de investigación desarrollados y los elementos del contexto que pudieron afectarlos, aspectos que se desarrollan más ampliamente en el siguiente capítulo. A partir de éstos procedió, con un enfoque multidisciplinario y cuanti-cualitativo para:

- a)* ofrecer una visión global de cómo se organiza y funciona la evaluación en el sector, en términos formales e informales,
- b)* metaevaluar cada uno de los programas evaluados,
- c)* realizar una síntesis de las anteriores como metaevaluación transversal del conjunto, y
- d)* analizar la evolución vivida por la evaluación de algunos programas sociales seleccionados.

En términos de las clasificaciones básicas revisadas, dicha aplicación se definió como una metaevaluación:

- a) sectorial, porque analiza las evaluaciones realizadas en el sector coordinado por la Secretaría de Desarrollo Social,
- b) formativa, ya que realiza su análisis con fines de perfeccionamiento futuro de esta práctica,
- c) *ex-post*, debido a que su valoración se realiza después de la aplicación de las evaluaciones,
- d) externa, por no tener vínculo directo con los operadores de los programas y haber realizado sólo una de las 31 evaluaciones consideradas en 2002, de poco peso en el conjunto, y
- e) integral, o sea, de diseño, proceso, resultados e impactos.

Esta metaevaluación utilizó el análisis documental (leyes, decretos, planes, programas, acuerdos, RO, TR, informes de evaluación, etcétera), pero también aplicó entrevistas abiertas y semiestructuradas a sus actores más relevantes (funcionarios, evaluadores, congresistas, organizaciones no gubernamentales, etcétera). En cambio, no hizo uso de técnicas propias de la hermenéutica como el análisis de textos.

Las evaluaciones fueron analizadas mediante un conjunto de indicadores o lista de comprobación que integran elementos de las distintas propuestas de criterios revisadas y los adaptan a nuestro contexto nacional:

- Decisiones y procesos del contexto inmediato a la evaluación de programas:
  - planeación de procesos de evaluación,
  - origen de la decisión de evaluar,
  - instancias intervinientes y su papel formal,
  - funcionamiento real del sistema: relaciones entre las instancias; licitaciones y TR; mercado de la evaluación; cumplimiento de obligaciones legales; equipos evaluadores y sus móviles; principales problemas administrativos,
  - recursos para evaluar,
  - difusión de los resultados de las evaluaciones, y
  - utilización de resultados en la toma de decisiones.

- Metodología y procesos internos a la evaluación de programas:
  - fuentes de información y análisis de datos,
  - pertinencia y diseño del programa,
  - recursos utilizados por el mismo,
  - análisis de procesos,
  - medición de cobertura,
  - resultados alcanzados,
  - impactos provocados,
  - participación y satisfacción social,
  - utilización de resultados y de recomendaciones previas,
  - valoración global del programa y recomendaciones,
  - escenarios prospectivos,
  - informes finales, y
  - costo de la evaluación.

Estos temas fueron valorados con base en los criterios generales de investigación (objetividad, validez, confiabilidad, representatividad, consistencia, etcétera) y con algunos específicos de la metaevaluación como pertinencia, autonomía, credibilidad, utilidad, participación y calidad.

A su vez, para la medición de la calidad se tomaron en cuenta tres elementos que se consideran relevantes para el contexto nacional: 1) contar con un diseño, recolección y tratamiento de datos rigurosos, 2) cumplir con los requerimientos especificados en el Acuerdo y los TR, y 3) concluir con la explicitación de un juicio valorativo fundamentado sobre el programa.

Dando continuidad al enfoque anterior, hoy disponemos en México del trabajo desarrollado por Vargas (2007) a las evaluaciones realizadas desde 2001 en adelante a tres programas de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Sagarpa): Alianza para el campo, Apoyo a excedentes comercializables y Apoyos directos para el campo. El trabajo se ubica en el marco del NPM y define a la metaevaluación como “la reflexión sobre el enfoque y el método empleados, de sus resultados y sus recomendaciones” (p. 13). Sin embargo, desde el título mismo del trabajo no especifica con claridad que la metaevaluación analiza evaluaciones y no programas, lo que se vuelve a confundir al realizar un análisis lógico del conjunto de los programas. Al ser un trabajo contratado por el CEDRSSA de la Cámara de Diputados, presenta

la originalidad de enfatizar el papel que desempeña el Poder Legislativo en la valoración de los programas federales y la contribución que realizan las evaluaciones al cumplimiento de sus atribuciones.

En el mismo año, encontramos otra metaevaluación (Schwentenius y otros, 2007) aplicada a tres evaluaciones realizadas en 1998, 2001 y 2003 del programa de pagos directos Procampo, también a cargo de la Sagarpa. En ésta se recuperan las definiciones de metaevaluación de diversos organismos internacionales como la OCDE, el BM o la CEPAL,<sup>7</sup> en que se resalta como objetivo el análisis de la calidad de la evaluación y el desempeño de quienes la realizan. Sin embargo, en otro apartado se plantea como objetivo identificar “qué diferencia ha marcado el programa en comparación con los no beneficiarios y qué impacto ha tenido sobre los beneficiarios” (p. 110), lo que correspondería más bien a un diseño experimental de evaluación de impacto, que suponemos es lo que consideran debería haberse realizado para que se establecieran relaciones de causalidad entre el programa y los aumentos de ingresos de los beneficiarios.

Realiza un análisis crítico de las evaluaciones con la intención de cuestionar el supuesto resultado exitoso en términos de impacto y efectos multiplicadores que éstas adjudican al programa. Identifica inconsistencias en los objetivos, metodología, interpretación de resultados y conclusiones, que limitan la calidad y utilidad de dichas evaluaciones, como por ejemplo, “resultados derivados básicamente de las percepciones de los beneficiarios encuestados y del análisis lineal de las respuestas” (p. 113), ambigüedades del cuestionario y forma en que se levantaron las encuestas o parcialidad en las valoraciones como consecuencia de la falta de libertad de los evaluadores. Concluye que las evaluaciones no aportan evidencia suficiente para sostener que los beneficiarios han logrado mejorar su calidad de vida.

Otro trabajo que, si bien no menciona el término de metaevaluación, alude al mismo, es el aplicado a la evaluación del programa Oportunidades (Sedesol, 2008), publicado en la página electrónica del BM como parte del material utilizado en el Taller de análisis y reflexión de programas de transferencias condicionadas. El mismo cubre las evaluaciones realizadas en 1997-2006 y

<sup>7</sup> No aparecen referencias a trabajos de metaevaluación realizados en otros contextos nacionales, salvo el de Larrú en la Unión Europea.

subraya su marco normativo; la importancia del modelo aplicado, sus resultados y utilidad; así como sus fortalezas y retos. Bajo el título de *Impactos* se presentan resultados que no siempre lo son, como por ejemplo el número de consultas atendidas. Se confirma el notable incremento en los presupuestos destinados a las evaluaciones que ya se había mencionado en Cardozo (2006a): se ha más que triplicado en el periodo 2001-2007. Sin embargo, el trabajo no constituye un análisis crítico sino un intento por resaltar las supuestas bondades del programa, sin mencionar debilidad alguna, que fue realizado por la principal institución que ha tenido a su cargo las evaluaciones en estudio (Cardozo, 2009).

Una nueva metaevaluación de las 16 evaluaciones realizadas en 2006 a la gestión de 2005 y de su evolución en 2001-2006 en el campo del desarrollo social federal fue realizada con una lista de chequeo similar a la utilizada en los trabajos de 2002-2003 para lograr su comparación. Se trata de la misma lista en los aspectos relativos a la metodología aplicada por los evaluadores y otra cercana en materia de organización del sistema, ya que revisa sólo los aspectos en que se presentaron cambios posteriores al primer trabajo: normatividad, mecanismos de selección, contratación y pagos, recursos utilizados, difusión y utilización de resultados, institucionalización de la cultura de evaluación y principales problemas detectados (Cardozo, 2009). Nuevamente, sus principales resultados se sintetizan en el siguiente capítulo.

En 2010, se realizó la primera metaevaluación de las evaluaciones realizadas en 2008-2009 en el campo del desarrollo social a cargo del GDF (Cardozo y otros, junio de 2010). Evalúa DF se interesó en evaluar su propio proceso, los métodos y la calidad de los trabajos de evaluación realizados, por lo que la metaevaluación incluyó la revisión de diseño, procesos, productos e impactos alcanzados. Se revisaron múltiples variables que afectan el nivel de calidad alcanzado por una evaluación, se hizo su medición, identificando problemas y proponiendo posibles soluciones. También en este caso, el contenido del trabajo realizado se sintetiza en el capítulo siguiente.

Debido a que el trabajo de metaevaluación anterior podía ser parcialmente cuestionable por haber sido coordinado directamente por una de las consejeras de Evalúa DF, en 2010 se decidió contratar por concurso una segunda metaevaluación que cubriera las cuatro evaluaciones de políticas realizadas en 2009-2010, que fue desarrollada por investigadores de la Universidad de Sevilla, la Universidad Carlos III de Madrid y la Sociedad Española de Evaluación.

Dicho trabajo se concentró, en forma similar al anterior, en tres apartados: 1) el diseño de la metaevaluación, su marco teórico, contexto, objetivos y enfoque metodológico; 2) el marco legal, la organización y el funcionamiento del sistema de evaluaciones externas, explorando sus principales hitos, fases y etapas a partir de las funciones que desempeñan los principales actores del sistema y examinando su incidencia sobre los procesos de evaluación llevados a cabo, y 3) el análisis de las políticas evaluadas en 2009 con base en los cinco grandes criterios/dimensiones en los que se vertebró el enfoque metodológico: contexto y encargo de las evaluaciones, medios del sistema, calidad técnica, utilidad y ética (Osuna y otros, 2011).

Como consecuencia de las dos metaevaluaciones realizadas en Evalúa DF, se han identificado algunos aspectos organizacionales y metodológicos cuyo perfeccionamiento se espera contribuya a lograr mejores evaluaciones e, indirectamente, al rediseño y operación de políticas y programas sociales.

Finalmente, las metaevaluaciones mexicanas también identifican importantes problemas en la utilización de las recomendaciones surgidas de los procesos de evaluación. A nivel federal, en que las recomendaciones no son vinculatorias, Sedesol contaba con un Sistema de seguimiento a evaluaciones externas (Sisee) y vigilancia estrecha de los compromisos asumidos y sus tiempos; pero, dicho mecanismo parece haber sido efectivo únicamente al inicio de su aplicación, ya que poco a poco los funcionarios fueron disminuyendo la magnitud de sus compromisos en los convenios correspondientes o se limitaron a hacerlos meramente formales (ej. envío de un oficio solicitando algo acordado) ante el temor de que la SFP o la ASF pudieran acusarlos de incumplimiento. Por otra parte, el Coneval y la SHCP diseñaron un sistema computarizado de control de gestión que permitiría visualizar fácilmente los puntos fuertes y las debilidades de cada uno de los aspectos evaluados a los programas (diseño, operación, impactos, etcétera), similar a un tablero de control o al barómetro utilizado en Francia por Ohana.

A nivel de la SFP, la opinión es que existe un avance en la utilización de recomendaciones pero aún muy insuficiente, por lo que a partir de los Programas de mejoramiento de la gestión se intentaría introducir un nuevo mecanismo de seguimiento de recomendaciones. La idea seguiría siendo la que instrumentaba Sedesol: que los operadores expresen cómo tendrán en cuenta las recomendaciones y en qué plazo, o bien que argumenten las razones que

no hacen factible su aplicación y propongan una solución diferente. También podrían fundamentar su rechazo no sólo a la solución propuesta, sino al problema mismo identificado. Actualmente Coneval informa sobre los convenios de recomendaciones y sus resultados.

Por otro lado, no se perciben avances en la utilización de las evaluaciones por los diputados federales. Al ser entrevistados algunos demuestran no saber ni siquiera que existen evaluaciones externas y confundirlas con las auditorías realizadas por la ASF (Cardozo, 2009).

En el caso del GDF la situación es diferente porque, con base en las recomendaciones emitidas en las evaluaciones, Evalúa DF elabora las suyas propias y éstas tienen carácter vinculatorio; sin embargo, se están enfrentando dos tipos de problemas: recomendaciones que no fueron aceptadas por los órganos que operan las políticas y programas evaluados y cuya controversia debería ser resuelta por una comisión que lleva meses sin reunirse; y recomendaciones que, si bien fueron aceptadas, no se están cumpliendo plenamente, en cuyo caso Evalúa DF no tiene más camino que informar a la Contraloría del DF para lo que proceda.

## **Comentarios finales**

El capítulo sintetiza las dificultades culturales, sociopolíticas, administrativas y metodológicas que enfrenta el desarrollo de procesos generalizados de evaluación y los aportes que respecto de su calidad pueden realizar los trabajos de metaevaluación, dirigidos a enjuiciar el valor y el mérito de los trabajos de evaluación, tanto en términos metodológicos como organizacionales.

Se presentan los limitados avances teóricos, metodológicos y empíricos producidos por distintos países en materia de metaevaluación, y se menciona la localización de un solo trabajo de este tipo en América Latina y al margen de México, resaltando la poca utilización de sus resultados en todos los ámbitos. Luego se describen en términos generales los diversos trabajos realizados en México en los últimos cinco años, así como los esfuerzos emprendidos, a nivel federal y del GDF, para lidiar con el problema de su insuficiente utilización en los procesos de rediseño e implementación de políticas y programas.

Con la intención de profundizar en materia de metaevaluaciones mexicanas, en el siguiente capítulo se presentan algunos ejemplos de aplicación.



## CAPÍTULO VIII

### Ejemplos de metaevaluaciones realizadas en México

Una vez revisada la teoría, métodos y experiencias relativas a la metaevaluación, el presente capítulo se propone brindar algunos de los ejemplos de trabajos que se han realizado en México en los últimos años.

Con este objetivo, se presenta una síntesis de lo realizado en los trabajos dirigidos a valorar la calidad de las evaluaciones efectuadas y sus factores determinantes en programas de desarrollo social de los siguientes casos: evaluaciones federales de 2002-2003, evaluaciones federales de 2005-2006, y evaluaciones del GDF de 2008-2009.

Las mismas han sido desarrolladas conforme al enfoque propuesto en el capítulo anterior: acercamiento interdisciplinario integral, utilización de listas de comprobación que cubren los aspectos organizacionales y metodológicos, aplicación de criterios generales y específicos, elaboración de índices de calidad, identificación de problemas y sugerencia de posibles soluciones para mejorar los trabajos de evaluación, e, indirectamente, el rediseño y la puesta en práctica de políticas y programas.

#### **Metaevaluación de programas federales de 2002-2003**

Sintéticamente, la metaevaluación a que se refiere el título (Cardozo, 2006a) arribó a la siguiente caracterización del sistema de evaluación, los métodos empleados por los evaluadores y la calidad de sus 31 productos:

- Exigencias de la Cámara de Diputados de muy difícil cumplimiento: evaluar todos los programas sujetos a RO, todos los años y en todos sus aspectos.
- Poca pertinencia de la legislación vigente utilizada, falta de criterios claros para decidir la forma de contratación de las evaluaciones y ausencia de los mismos para determinar techos financieros.
- Imprecisiones y ausencias en las solicitudes de evaluación (bases y TR), debidas, en parte, a poca claridad del Acuerdo de 2002.
- Altas expectativas de Sedesol respecto de la posibilidad y conveniencia de aplicar estudios experimentales.
- Existencia de importantes limitaciones de los recursos institucionales (presupuestales y de información) para el desarrollo de trabajos de calidad, con excepciones relevantes (Oportunidades) y ausencia de un mecanismo de distribución racional de los mismos.
- Insuficiente formación o capacitación en materia de evaluación de algunos funcionarios responsables del proceso y de diputados encargados de tomar en cuenta los resultados incluidos en los informes de evaluación.
- Conformación en proceso de un mercado de evaluación que contaba con una oferta restringida de capacidades especializadas de los equipos evaluadores en el tema, agudizada por la falta de programas universitarios de formación y capacitación en la materia, de disposición de guías de evaluación y de asociaciones de profesionales especializados.
- Falta de autonomía financiera e institucional suficiente del evaluador (contratado y pagado por el propio programa que evalúa) para garantizar su objetividad.
- Problemas de suficiencia, oportunidad, consistencia y confiabilidad de la información institucional disponible.
- Aplicación, con pocas excepciones, de instrumentos a muestras de representatividad desconocida o dudosa: determinación de tamaños muestrales, mecanismos de selección, e inclusión de dicha información en el informe para que el lector pueda valorar la calidad de los datos en que se basa el trabajo realizado.
- Casi nula participación social (a lo sumo limitada a la respuesta de encuestas por los beneficiarios) en los procesos de gestión y evaluación de los programas (en contradicción con la intención manifestada por planes y programas

gubernamentales), debido que se tiende a priorizar la *expertise* económica y econométrica del evaluador.

- Cobertura parcial de los aspectos solicitados y otros de reconocida relevancia para arribar a conclusiones bien fundamentadas.
- Emisión de pocos juicios explícitos de valoración fundamentada del programa analizado.
- Restringida difusión y utilización de sus resultados para el mejoramiento de los programas y la asignación de recursos, en parte, debido al carácter no vinculatorio de sus recomendaciones.
- Desarrollo limitado de una cultura de la evaluación que valore su contribución, cambie actitudes y se traduzca en nuevas prácticas que integren la evaluación a la planeación y la operación.

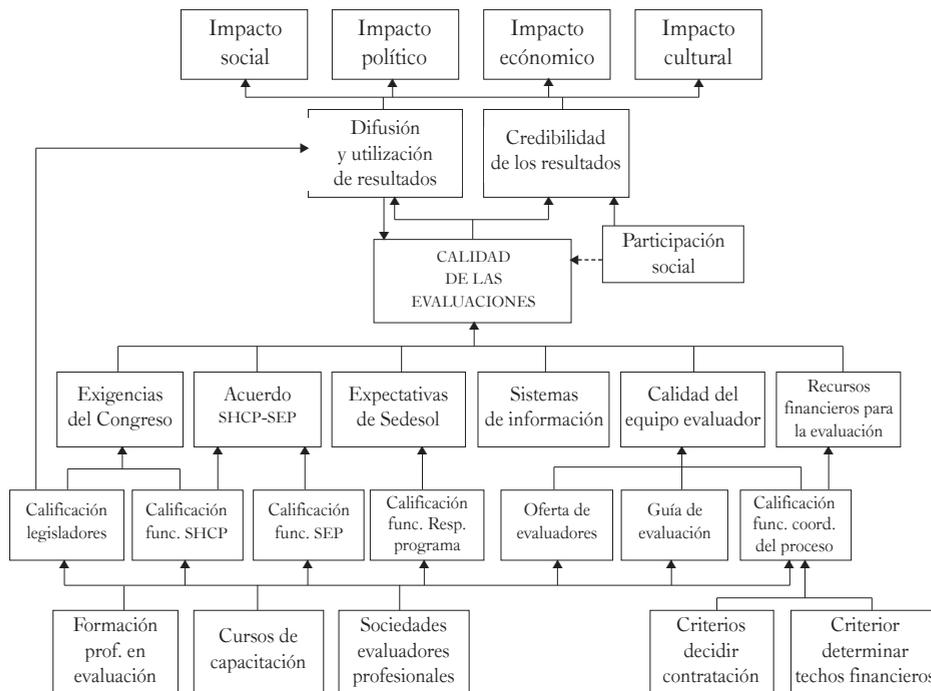
En cuanto a la calidad de lo realizado, ésta se midió con base en la coherencia lógica entre los cuatro elementos identificados por Patton (1997) en toda evaluación: datos, interpretaciones, valoraciones y recomendaciones; y en uno de los métodos propuestos por Abate (2000), consistente en una combinación de tres indicadores: diseño, recolección y tratamiento riguroso de datos; cobertura de los aspectos revisados en la metaevaluación; y emisión de un juicio global y fundamentado de valor sobre la gestión evaluada.

Complementariamente, se aplicó otro de los métodos sugeridos por Abate, que implica recoger la percepción sobre el tema de los diferentes actores vinculados al proceso, lo que se realizó mediante la aplicación de 25 entrevistas. Ellos reconocieron que el avance de la evaluación en México aún es limitado e identificaron como las principales causas de esta situación las culturales, sociales y políticas; en cambio, dieron un peso mucho menor a los problemas metodológicos que se relacionan en gran parte con la calidad de las evaluaciones analizadas.

Como consecuencia de lo anterior, se identificaron 16 evaluaciones de baja calidad, siete de calidad regular, siete de calidad alta y ninguna excelente. Ante los problemas detectados (Diagrama 5) se propusieron: impulsar la formación profesional y los cursos de capacitación en evaluación, conformar sociedades de evaluadores profesionales, establecer criterios legales para decidir sobre las formas de contratación y para determinar techos financieros para la evaluación, mejorar los sistemas de información institucional, impulsar una verdadera participación social en la gestión y evaluación, publicar y difundir ampliamente

los resultados de evaluación, crear a mediano plazo un organismo autónomo responsable de las evaluaciones y promover el desarrollo de una cultura y de prácticas efectivas de utilización de la evaluación.

DIAGRAMA 5  
*Red de variables implicadas en los problemas de evaluación federal*



Fuente: elaboración propia.

También se analizó la evolución de las evaluaciones realizadas en tres programas seleccionados,<sup>1</sup> sin pretensión de representatividad, arribando a la conclusión de que mostraban una peligrosa tendencia a la disminución de la calidad del trabajo efectuado.

<sup>1</sup> Progres-Oportunidades 1997-2003, Microrregiones-Desarrollo Local 2002-2003 y Promoción de Culturas Indígenas-Desarrollo de Culturas Indígenas 2001-2003.

La evaluación federal se ha interesado siempre en la búsqueda de los métodos más rigurosos y objetivos, lo que la ha llevado a preferir las mediciones cuantitativas y las mediciones de impacto mediante métodos experimentales. En la práctica, estos últimos han sido poco utilizados por falta de grupo de control, de información sobre la línea de base del programa, o de presupuesto para cubrir su costo.

En el informe del BM (2004) se confirmaron algunos problemas y soluciones expuestos en este libro. Se reconoció el esfuerzo realizado por el gobierno federal mexicano, especialmente la Sedesol, en materia de monitoreo y evaluación de los programas dirigidos a la lucha contra la pobreza.<sup>2</sup> Pero, identificó una agenda pendiente en materia social: integración más eficaz de la política económica y la social, calidad de los servicios, acceso a seguridad social, políticas para el incremento de ingresos y la mejora de su distribución (especialmente en el caso de grupos indígenas), mejora en la productividad de los autoempleados y de las empresas informales, fortalecimiento de la rendición de cuentas, la participación social y la administración basada en resultados (ABR), a partir de un enfoque integrado de M&E, como el que estaba financiando en Argentina.<sup>3</sup>

Opinaba que la estrategia Contigo, iniciada en el año 2000, seguía siendo difícil de evaluar y que no quedaba clara la relación entre su gran número de programas y el enfoque integral perseguido por el gobierno (variaciones en la focalización de acuerdo con la Secretaría que lo implementa, por ejemplo). Agregaba que se sabía poco respecto del gasto descentralizado en combate a la pobreza y que no era evidente el incentivo que induciría a los estados a interesarse en su aplicación.

Consideraba excelente la evaluación de impactos iniciada a Progresá en 1998, en la que ponía más atención que en la que, en ese momento, se aplicaba a Oportunidades. Sin embargo, opinaba que se requería comprender mejor los determinantes de la cobertura y la calidad de los servicios públicos dirigidos a

<sup>2</sup> La reducción de la pobreza por ingresos mostraría resultados estadísticamente significativos a nivel rural y urbano; en cambio, la pobreza estimada en términos de consumo sólo los exhibiría en el ámbito rural.

<sup>3</sup> También en Sedesol se estaba proponiendo que la dirección encargada de la evaluación pasara a llamarse Dirección General de Monitoreo y Evaluación de Programas Sociales.

la población pobre, así como a las condiciones institucionales que afectaban a la incorporación social y la rendición de cuentas.

El diseño experimental, si bien lo consideraba la técnica más novedosa de medición de impacto, opinaba que es “más limitada para el aprendizaje continuo de los procesos que están detrás de los cambios observados y para comprender el papel del contexto local”, así como “no puede controlar la influencia del tiempo en los resultados que no se relacionan con el programa” (BM, 2004:LVII y 242), los que resultan decisivos para mejorar el diseño de los programas. Pensaba que eran los sistemas de monitoreo con base en el seguimiento continuo de indicadores los que permitirían mayor aprendizaje y participación social, razón por la que proponía su combinación con las evaluaciones en profundidad, tema en el que la Sedesol continuó trabajando.

También concordaba con las conclusiones de este trabajo en el sentido de que los programas requieren tiempo para que se evalúen sus impactos, lo que no puede hacerse todos los años, sino en forma selectiva. Mencionaba que la realización de tantas evaluaciones de impacto llevaba a que éstas fueran rápidas, superficiales y arrojaran pocos beneficios, por lo que recomendaba su complementación con un sistema de monitoreo como el que se estaba preparando en la Oficina de la Presidencia, que permitiría una cierta anticipación de impactos. Además, sugería interacciones entre evaluadores internos y externos, con énfasis en la participación de los beneficiarios. Consideraba que sólo tiene sentido valorar la efectividad de los programas sociales que fueran innovadores y reproducibles, y tuvieran una alta cobertura nacional, intervenciones bien definidas y que absorbieran recursos sustanciales (BM, 2004).

El BM opinaba también que la LGDS enfrentaba desafíos para su implementación debido a la rigidez que implicaba en material presupuestal, las funciones diseñadas para las comisiones creadas y el énfasis puesto en la gestión de los municipios.

El documento concluía que: 1) la Sedesol debía expandir en forma gradual pero continua su sistema de monitoreo y evaluación, tomando más en cuenta las necesidades de información de la propia Secretaría, 2) el sistema requería claridad de metas y un alineamiento de todos los funcionarios trabajando en ese sentido, lo que el BM no percibía en 2004, y 3) era necesario redefinir el mandato del Congreso para que no exigiera evaluaciones anuales de impacto, sino un monitoreo del desempeño, recurriendo a una mezcla de evaluaciones internas y externas.

Todo lo anterior consideraba que debía complementarse con un mayor trabajo en temas de liderazgo, en el diseño de incentivos (sanciones y recompensas) y en la definición de funciones y responsabilidades.

### **Metaevaluación de programas federales de 2005-2006**

En el marco de un importante retraso en la expedición del Reglamento de la Ley General de Desarrollo Social y en la integración del Consejo Nacional de Evaluación, se llevaron a cabo las 16 evaluaciones de 2005-2006 que, como ya se mencionó, debían incluir la evaluación del 2005 y un metanálisis de 2002-2005.

Retomando a Glass (1976) “el metaanálisis consiste en un procedimiento estadístico para comparar los hallazgos de diferentes estudios que confluyen en un mismo objetivo, identificar tendencias y realizar generalizaciones” (en Cardozo, 2006a:138). Sin embargo, en el caso que se analiza no se trató de un procedimiento estadístico, pero sí de un análisis comparativo de evaluaciones de diferentes años que confluyen sobre el mismo programa. En parte tiene también elementos de la síntesis evaluativa que Bustelo (2002) define como “resumen de los resultados de múltiples evaluaciones efectuadas a políticas y programas similares” (citado en Cardozo, 2006a:137). El resultado del trabajo realizado fue, en realidad, una comparación de los resultados 2002-2004, en que prácticamente no se evaluó la gestión 2005 y, en consecuencia, no entró en la comparación realizada.

Respecto del diagnóstico del sistema de evaluaciones externas elaborado en 2002-2003, se identificaron pocos cambios, por lo que el mismo se mantuvo en 2006, con algunos comentarios adicionales:

- La LGDS, expedida en 2004, mantenía en general las exigencias previas de la Cámara de Diputados y aunque abría la posibilidad de flexibilizar algunas cuestiones como la evaluación de impacto, aún no estaba reglamentada.
- Los cambios experimentados en la Ley de Adquisiciones, aunque insuficientes, mejoraban un poco las condiciones para la contratación de evaluadores, pero la ausencia de criterios para determinar techos financieros continuaba igual. Como se acaba de mencionar, aún no se había cumplido con la obligación de expedir el Reglamento de la LGDS y de integrar el Consejo Nacional de Evaluación, que ya llevaban un importante retraso.

- En cuanto a las imprecisiones y ausencias por parte de la demanda, si bien hubo intentos de la SFP por mejorar el Acuerdo de 2002, éste seguía vigente, manteniendo los problemas ya detectados. Alguna mejoría podía percibirse en los TR, también por la mayor experiencia de los funcionarios encargados de redactarlos.
- Se mantenían las altas expectativas de Sedesol respecto de la posibilidad y conveniencia de aplicar estudios experimentales, a pesar de su poca factibilidad demostrada (en 2006 sólo una evaluación los utilizó).
- Se identificaron similares problemas de información y un acceso aún más limitado de recursos financieros disponibles para las evaluaciones. Además, en la medida en que se asignaba presupuesto mediante decisiones aisladas de los programas, la ausencia de un mecanismo de distribución racional de los recursos seguía vigente.
- Los cambios en torno a la calificación de funcionarios responsables del proceso y de diputados ocurrían en casos aislados que se interesaban por tomar cursos y estudiar por su cuenta, pero en términos agregados resultaban marginales.
- En materia de capacidades especializadas en el tema, la situación había evolucionado muy poco, algunas capacidades adicionales se habían logrado a partir de la experiencia de ir “haciendo” evaluaciones sobre la marcha, pero la ausencia de programas formales de capacitación se mantenía casi igual.
- La falta de autonomía financiera e institucional suficiente del evaluador para garantizar su objetividad seguía vigente con los mismos riesgos.
- Los problemas de suficiencia, oportunidad, consistencia y confiabilidad de la información institucional disponible no presentaban mejorías evidentes.
- La aplicación de instrumentos a muestras de representatividad desconocida o dudosa, se mantenía con pocas excepciones. Si bien era común encontrar un anexo metodológico para la medición de cobertura en las propuestas técnicas, éste a menudo no se aplicaba en el desarrollo del trabajo, y casi ningún evaluador informaba sobre sus tamaños muestrales y mecanismos de selección.
- Se habían creado algunos comités y contralorías sociales para hacer avanzar la participación social, pero no funcionaban en la práctica.
- En relación con la cobertura de los aspectos solicitados en la evaluación, el descuido de aspectos como participación social e impacto, o el inadecuado tratamiento de temas como focalización y cobertura, seguían vigentes.

- En lo que respecta a la emisión de juicios explícitos de valoración fundamentada del programa analizado, éstos seguían siendo muy escasos y a lo sumo se identificaban fortalezas, desafíos y recomendaciones que suponían un juicio implícito de que se justificaba la continuidad del programa.
- La difusión iba mejorando poco a poco, ya era más común encontrar los informes de evaluación en internet, aunque este mecanismo tenía obvias restricciones debido a que la mitad de la población no tenía acceso a la red, especialmente las personas de menores recursos a los que iban dirigidos los programas. La utilización de resultados seguía siendo muy limitada, pero la Sedesol intentaba llegar a acuerdos con los operadores de programas para que las recomendaciones se tomaran en cuenta.
- Continuaba el desarrollo limitado de la cultura de la evaluación, al igual que en 2005.

En cuanto a la calidad de lo realizado, ésta se midió con base en la misma metodología aplicada en la metaevaluación anterior, y su resultado mostró que ninguna evaluación presentaba una excelente cobertura de todos los aspectos considerados en el trabajo como relevantes, seis eran de calidad alta, ocho de calidad regular y en el extremo opuesto, se identificaron dos evaluaciones de muy pobre calidad. Esto implicaba una sensible mejoría respecto de 2002 (disminuyen notoriamente las malas evaluaciones y los resultados se trasladan a los otros dos tipos: regular y alta), mediante las 11 entrevistas complementarias aplicadas se pudo confirmar que, aunque la calidad se consideraba variable, en promedio se la juzgaba regular.

Debido a que, a pesar de que se solicitó al Instituto Federal de Acceso a la Información Pública (IFAI), la información correspondiente a los costos y mecanismos de asignación de las evaluaciones realizadas, sólo se recibió una respuesta parcial, fue imposible analizar su relación calidad/costo, como se había hecho en el ejercicio de metaevaluación precedente. Con los datos entregados por Sedesol para la mitad de las evaluaciones realizadas, sólo se puede afirmar que la más cara de ese subgrupo fue efectivamente la de mejor calidad, mientras la segunda más cara no se destacó en la otra variable en estudio.

*En conclusión*, en 2006 se percibe una mejoría aún insuficiente de la calidad de las evaluaciones y la vigencia de prácticamente todos los problemas identificados

tres años antes, por lo que se mantuvieron las recomendaciones del trabajo previo (Cardozo, 2009).

La creación del Coneval en 2006, aunque se ha caracterizado por la continuidad metodológica de fondo, ha resuelto algunos problemas importantes como el contar con lineamientos más claros para la ejecución del trabajo, incluir en forma expresa la evaluación del diseño del programa (originalmente referida sólo a su operación), o solicitar su reestructuración con base en el marco lógico.

### **Metaevaluación de programas del GDF de 2008-2009**

Como ya se mencionó, Evalúa realizó en 2010 su primera metaevaluación a las evaluaciones desarrolladas en 2008-2009, incluyendo la revisión del diseño, los procesos de puesta en práctica, los productos generados y su impacto.

La forma en que se organiza y funciona la evaluación, así como sus métodos, en el caso del GDF responden a una práctica y un enfoque diferente (de tendencia constructivista) del sustentado por el Coneval (Cardozo, 2011).

Los resultados encontrados se presentan a continuación (Cardozo *et al.*, 2010):

- Marco teórico-conceptual de la metaevaluación: se llevó a cabo una revisión del estado del arte en la materia y se propuso realizar una metaevaluación integral y formativa. Integral, porque incluye todas las etapas del proceso de evaluaciones realizadas: diseño, instrumentación y resultados. Formativa, porque realiza su contribución con fines de perfeccionamiento futuro de esta práctica.
- Experiencias de metaevaluación: se mencionaron experiencias en organizaciones y países europeos y americanos, especialmente las desarrolladas en México en los últimos años.
- Políticas y programas de desarrollo social en el GDF: se describieron los objetivos, principios y características de la política social, los tipos de programas aplicados, así como los principales programas, las dependencias que los operan, sus presupuestos y metas, indicando cuáles de ellos fueron considerados en la evaluación de 2008.
- Marco normativo de la evaluación en el GDF: incluyó una revisión de los temas relacionados con la evaluación en la Ley de Desarrollo Social del

Distrito Federal y su Reglamento, la Ley Federal de Responsabilidades de los Servidores Públicos, la Ley Orgánica de la Administración Pública del DF, la Ley de Planeación del Desarrollo del DF, la Ley de Presupuesto y Gasto Eficiente del DF, el decreto del jefe de Gobierno para la creación del Consejo de Evaluación del Desarrollo Social del Distrito Federal y el Estatuto Orgánico de Evalúa DF. Se concluyó que se trata de un marco jurídico bastante completo aunque se señalan algunas incongruencias.

- Fuentes de información utilizadas y métodos aplicados por la presente metaevaluación: las técnicas de recolección y tratamiento de datos utilizados incluyeron, al igual que en los trabajos antes presentados, el análisis documental de leyes, decretos, planes, convocatorias, TR, propuestas técnicas, informes, recomendaciones, respuestas de los responsables de programas y controversias, con base en una lista de chequeo especialmente diseñada para el caso. Este se complementó con la aplicación de 16 entrevistas semiestructuradas a informantes-clave (Consejeros de Evalúa DF, evaluadores y responsables de programas), que permitieron ahondar en las percepciones, opiniones, perspectivas y propuestas de los diversos actores relacionados con la evaluación y contribuyeron a la detección de problemas y a la explicación y solución de los mismos.
- Organización y gestión del sistema de evaluaciones externas: se revisaron los avances que se han realizado en materia de institucionalización y profesionalización de la evaluación, los principios básicos que han guiado el trabajo en Evalúa DF, la composición y relaciones entre las instancias participantes (especialmente Evalúa DF, evaluadores y evaluados), los criterios utilizados en la definición de la agenda de las evaluaciones 2008 de Evalúa DF y del tipo de trabajos a realizar (evaluación de diseño y de operación de programas), los procedimientos aplicados por Evalúa DF (elaboración y publicación de la convocatoria y los TR correspondientes, adjudicación de contratos, acompañamiento, valoración de informes, emisión de dictámenes, elaboración de recomendaciones y controversias), los recursos financieros disponibles y su forma de administración (fondo independiente de los programas), las características de los recursos humanos de Evalúa DF, de los operadores de programas que pudieron entrevistarse, y muy particularmente, de los equipos evaluadores (grado de estudios, especialización pertinente al programa, interdisciplinariedad, experiencia previa).

Los principales problemas que los actores entrevistados identificaron en materia de organización y gestión del sistema de evaluaciones externas fueron: temor de los operadores ante la evaluación y sus consecuencias por falta de suficiente comunicación del tipo de evaluación y las actividades de Evalúa DF, que se tradujo en insuficiente apertura y disposición del personal de algunos programas para entregar la información documental disponible y acceder a entrevistas.

- Análisis de las evaluaciones externas, en las que se detectó: 1) un cumplimiento parcial de los contenidos solicitados en los informes finales, 2) una confusión e insuficiente inclusión de los elementos que componen un marco teórico-conceptual, 3) una adecuada utilización de las fuentes de información aunque sería conveniente que se ampliara la consulta de publicaciones académicas pertinentes al tema, 4) se hace referencia sólo a los métodos de recolección de información y una descripción de actividades realizadas, omitiendo ampliamente referirse a métodos de análisis y síntesis, 5) la evaluación del diseño del programa se puede calificar como suficiente aunque presenta muchos problemas en dos casos de programas centrales y resulta claro que las evaluaciones delegacionales han tenido muchas más dificultades para lograr buenos resultados, 6) la evaluación de operación resulta más débil que la de diseño por falta de tiempo para cubrir el tema a la fecha de entrega del informe y/o necesidad de contar con información de campo (observación directa, encuestas, entrevistas), entre otras razones, 7) sólo excepcionalmente resulta aceptable la valoración del diseño, la operación y su congruencia, ya que se percibe resistencia de los evaluadores a realizarla o a fundamentarla en los hallazgos previamente realizados, 8) en la mayoría de los casos se identifican los problemas, pero falta el análisis de sus interrelaciones y prioridades, 9) las recomendaciones tienden a ser muy generales, no se propone cerrar programas aunque en algunos casos se hubiera justificado, y se descuida la sugerencia de procedimientos, plazos y autoridades a quienes deben dirigirse.

A pesar de ello, la mayoría de los entrevistados (evaluadores y Evalúa DF) las consideraron aceptables, aunque en varios casos quedan dudas sobre la satisfacción de las expectativas de los funcionarios de programas.

- Resultados de las evaluaciones externas: se revisaron los logros y desafíos pendientes en torno a los dos objetivos fundamentales de la evaluación: 1)

el mejoramiento de los programas para lograr una efectiva solución de los problemas sociales y el cumplimiento de derechos, y 2) los procesos de difusión, transparencia y rendición de cuentas que contribuyen a generar ciudadanía y hacer posible la participación.

El primero se analizó con base en las respuestas de los responsables de programas a los informes de evaluación, las recomendaciones de Evalúa DF, su nivel de aceptación, y el seguimiento de su efectivo cumplimiento. Se concluyó que hubo un buen nivel de respuestas a informes, tomando en cuenta que en las delegaciones se estaba terminando el periodo administrativo, y que sólo una de las evaluaciones centrales fue de fuerte rechazo al trabajo, las recomendaciones se concentraron en los programas centrales, se distribuyeron casi por igual entre diseño y operación, y 17% tuvieron un fuerte impacto presupuestal. En su conjunto, los resultados fueron satisfactorios, faltando aún los procesos de controversias y un seguimiento más efectivo y sustancial de las recomendaciones aceptadas.

En cuanto al segundo aspecto, se revisaron las tres solicitudes de información presentadas ante la Oficina de Información Pública y la difusión realizada a partir de reuniones de trabajo, buzón especialmente diseñado, conferencias de prensa y el portal de internet, concluyendo que se debería enfatizar la publicación de resultados por otras vías más accesibles a toda la población (periódicos, folletos, etcétera).

- Conclusiones sobre la calidad y utilidad de las evaluaciones externas realizadas: se revisó un conjunto de variables que afectan la calidad de una evaluación como su marco normativo, el funcionamiento del órgano encargado de su diseño, análisis y selección de propuestas, contrataciones, acompañamiento y dictaminaciones, la actitud de los funcionarios responsables de los programas, las características del equipo evaluador, y la relación entre los distintos actores institucionales vinculados con la evaluación; concluyendo que su desempeño es en general bueno, aunque obviamente mejorable.

Por su parte, la calidad misma del proceso de evaluación y su producto se midió a partir de tres indicadores: 1) marco teórico-conceptual, jurídico y de planeación global; 2) diseño metodológico de la recolección y el tratamiento de datos; y 3) la emisión expresa de un juicio valorativo, identificación y priorización de problemas y formulación de propuestas fundamentadas en el análisis del programa.

En relación con el primer indicador, se concluyó que los informes son extremadamente variados en su consideración y que todos dejan fuera algún aspecto importante, por lo que su tratamiento no se considera totalmente suficiente, salvo en dos casos que analizan los principales conceptos, enfoques, criterios de evaluación, y los planes del GDF. Respecto del segundo, se presentó un manejo aceptable de los métodos de recolección de información y una descripción de actividades realizadas, pero aparecen ausencias o insuficiencias generalizadas en materia de métodos de análisis y síntesis de la información. Finalmente, en torno al tercero, se encontró bastante reticencia de los evaluadores a efectuar las valoraciones demandadas del diseño, la operación y su congruencia. Los resultados son muy dispares: cinco evaluaciones se consideran de buen nivel, correspondiendo todas a programas centrales; mientras cuatro presentan algunas limitaciones importantes, dos delegacionales y dos centrales.

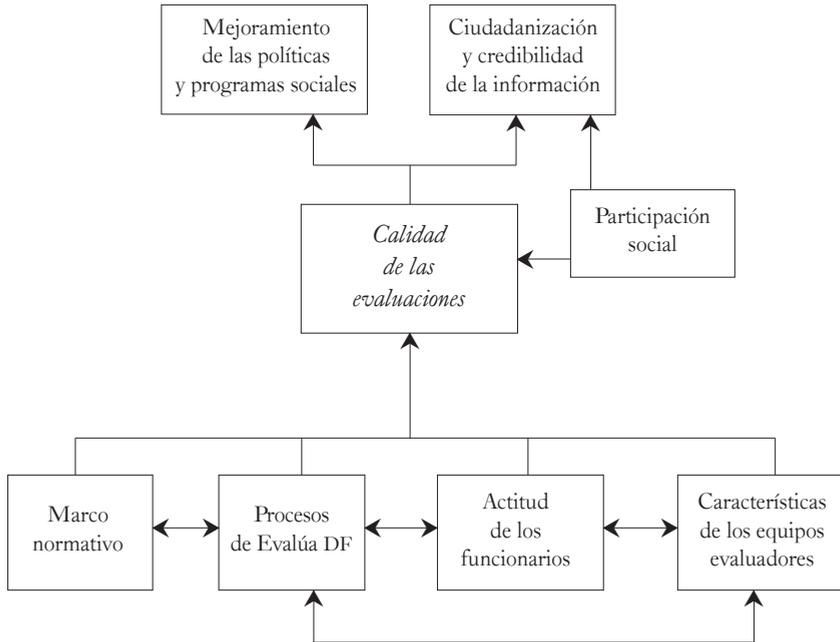
Con la intención de incidir en el mejoramiento de la calidad, la sugerencia lógica es poner mayor atención a sus factores determinantes, principalmente: marco normativo, procesos desarrollados por Evalúa DF, actitud de los funcionarios, características de los equipos evaluadores, y relación entre los tres principales actores del proceso (Diagrama 6).

Finalmente, se consideró que, a pesar de la buena respuesta de los programas a las evaluaciones y recomendaciones, aún no puede valorarse en forma definitiva su utilidad, dado que los procesos de cumplimiento o de controversias aún están en marcha. Ante la situación presentada, las principales propuestas se sintetizan por temas:

- Términos de referencia: definir claramente lo que se entiende por desarrollo social, particularmente la inclusión de la infraestructura social en el caso de los programas delegacionales; e incluir preferentemente temas más que preguntas exhaustivas, en todo caso cuidar de no ser repetitivos, formularlos en forma más específica en función de cada programa, y utilizarlos sólo como guía para la evaluación.
- Tiempos: ir recorriendo la publicación de convocatorias, en lo posible a los primeros meses del año, de manera de aprovechar mejor el tiempo para el desarrollo de las evaluaciones. También ampliar los plazos para que los funcionarios respondan a los informes de las evaluaciones y a las recomendaciones.

DIAGRAMA 6

*Árbol de relaciones entre variables de calidad y utilización de evaluación*



Fuente: elaboración propia.

- Marcos teóricos: precisar más lo que se espera en los TR para que las propuestas técnicas y los informes sean más completos y precisos.
- Métodos: incluir también mayor precisión, especialmente en materia de métodos de análisis y síntesis.
- Técnicas: ser más específicos al solicitar técnicas como el Marco Lógico y asegurar habilidades en el equipo evaluador que se consideren necesarias y hayan sido incluidas en los TR (manejo de la información estadística, elaboración de diagramas de flujo, etcétera).
- Acompañamiento: tratar de homogeneizar el trabajo realizado por lo Consejeros en cuanto a claridad y tiempo.
- Recursos: lograr que Evalúa DF cuente con un mayor presupuesto anual asegurado y una más amplia y más calificada planta laboral en el personal técnico-operativo, de manera de avanzar en la realización de evaluaciones

de impacto, sin detrimento de la calidad de su trabajo; y diversificar más los montos pagados en función de la complejidad de cada evaluación. En todo caso, reducir el número de evaluaciones anuales, pero no su alcance.

- Equipos evaluadores: mejorar la experiencia previa en trabajos de evaluación de programas sociales, asegurar mayor pertinencia de la interdisciplinariedad en función del programa evaluado y mayor conocimiento de la materia a ser evaluada.
- Información general: exigir una mayor revisión de la bibliografía académica publicada sobre cada tema evaluado.
- Información del programa: realizar una reunión previa al inicio de la evaluación para conocer mejor el funcionamiento del programa y sus principales problemas, así como los aspectos de mayor interés para los funcionarios que los operan, de manera de precisar y preparar la información requerida. También para sensibilizarlos sobre las atribuciones, objetivo y tipo de evaluación que Evalúa DF organiza.
- Diseño/operación: poner más atención a la operación y su congruencia con el diseño en los TR, menos tratada que esta última en los informes, y disponer de más tiempo y recursos para cubrirla mediante observación directa o encuestas a beneficiarios y derechohabientes.
- Participación social: generar mecanismos que promuevan la participación efectiva de los ciudadanos y funcionarios en los procesos evaluativos.
- Juicios valorativos: exigir en forma más contundente la valoración de los aspectos analizados con su correspondiente fundamentación en los resultados hallados.
- Propuestas: realizar sugerencias específicas para cada problema identificado, especialmente los catalogados como prioritarios, aprovechando las fortalezas del programa y cuidando de identificar la autoridad a la que deben dirigirse, los plazos y procedimientos de instrumentación recomendados.
- Información a evaluadores: brindarles información de los procesos posteriores a la entrega de su informe final: respuestas a éstos, emisión de recomendaciones, seguimiento de las mismas y, en su caso, controversias.
- Seguimiento de recomendaciones: si bien el porcentaje de recomendaciones aceptadas resulta satisfactorio, se requiere de un mayor seguimiento del cumplimiento de las mismas y la adopción inmediata de decisiones en caso de incumplimiento o falta de entrega de la información correspondiente.

- Desarrollo de la cultura de evaluación: a pesar de que se han hecho logros importantes en la construcción de esta cultura, aún se requiere mayor comunicación y coordinación entre los tres actores fundamentales de la evaluación. Para ello se recomienda promover la realización de cursos, diplomados, talleres, reuniones de discusión, participación en eventos académicos, publicaciones, entre otros.

### **Comentarios finales**

A nivel federal, las dos metaevaluaciones presentadas identificaron serios problemas, que caracterizaron al periodo 2001-2005, en el que los avances no fueron significativos: contar con un fondo independiente de recursos para la contratación de evaluadores al margen del programa evaluado, hacerlo siempre mediante licitación pública, avanzar en la realización de evaluaciones participativas centradas en el aprendizaje de funcionarios y ciudadanos, mejorar los esquemas de muestreo y su información, emitir valoraciones fundamentadas sobre la gestión del programa, realizar evaluaciones de impacto plurianuales y de conjuntos de programas, profesionalizar el campo, aplicar metaevaluaciones, y realizar un mayor esfuerzo por institucionalizar la cultura de la evaluación.

La constitución del Coneval en 2006 significaría un gran paso adelante en la mejor organización del sistema de evaluaciones externas y la calidad de los trabajos; sin embargo, se mantienen muchas de las debilidades y desafíos identificados en las metaevaluaciones anteriores.

La tercera metaevaluación, nos muestra que Evalúa DF optó por desarrollar un enfoque propio de la evaluación y una organización diferente de su sistema, lo que nos permite referirnos a él como un modelo emergente, que se distanció del que tradicionalmente se venía aplicando a nivel federal. Dada su experiencia, apenas iniciada a mediados de 2008, se considera que su desempeño ha sido satisfactorio, aunque se identifican múltiples variables en las que se puede mejorar su gestión.

Finalmente, dos comentarios que se consideran válidos tanto para el sistema federal como para el instrumentado en el GDF, a pesar de las diferencias existentes entre éstos:

1. Se percibe en ambos una urgente necesidad de fortalecer la utilización de los resultados de las evaluaciones, problema identificado también en las experiencias internacionales revisadas en el capítulo anterior. Se trata de que, a nivel federal, las dificultades surgen porque las recomendaciones no son vinculatorias y, a nivel del GDF, se están enfrentando problemas con los mecanismos previstos para hacer efectivo ese carácter.
2. Tanto Coneval como Evalúa DF deberían contar con mayor independencia de su respectivos gobiernos para llevar a cabo su importante función, en la misma forma que sucede con el IFAI o con el Instituto Federal Electoral.

## Consideraciones finales

Existe consenso sobre la necesidad de transformar a la administración pública tradicional para alcanzar de mejor manera los objetivos de las políticas y programas gubernamentales, pero no lo hay en cuanto a la dirección de dicho cambio, en el que se enfrentan quienes priorizan la solución de problemas técnicos y macroeconómicos apoyados en el NPM y quienes anteponen la solución de los conflictos y demandas sociales mediante el diálogo continuo entre gobierno y sociedad.

Esta disyuntiva se refleja en el terreno de la evaluación de programas sociales, también con dos tendencias claras: la aplicación de métodos experimentales y mediciones econométricas de trasfondo positivista, realizadas por especialistas; o, la utilización de métodos cualitativos que enfatizan la participación y el aprendizaje social, fundamentada en el constructivismo.

Si bien la literatura revisada muestra el énfasis experimental que la evaluación adquirió en Estados Unidos a fines de la década de 1960 y principios de la de 1970, todo parece indicar que su utilización ha disminuido sensiblemente debido a sus altos costos y discutibles resultados, lo que coincide con los comentarios críticos que el BM dirige al método. En cambio, la revisión hemerográfica en aquel país muestra que los estudios de carácter cualitativo están siendo más productivos y aceptados por los lectores.

A pesar de los debates teórico-conceptuales y metodológicos propios de una evaluación interdisciplinaria que hereda todas las discusiones presentes en las ciencias sociales y también aquellas que las separan de las ciencias físico-naturales, es posible afirmar que se cuenta actualmente con un marco bastante

sólido para su desarrollo, mismo que no presenta mayores innovaciones respecto de la revisión publicada por Cardozo (2006a).

A partir de la década de 1980 se ha incrementado sustantivamente el número de países con experiencias de evaluación, por lo que en el presente libro sólo se han presentado algunas que se han considerado podrían ser de utilidad en el proceso de aprendizaje mexicano. En esta edición, debido al giro tomado, se ha realizado una actualización del caso francés, y también se han incluido avances del conjunto de países integrantes de la UE, especialmente las recientes innovaciones españolas. A nivel latinoamericano, se mencionan los avances realizados en diversos países, como Brasil, Argentina, y la importante contribución debida a la conformación de una red de evaluadores, así como sus principales actividades.

A los pocos programas de formación y capacitación existentes en materia de evaluación se agregan algunos esfuerzos recientes, mexicanos y extranjeros, para mejorar la profesionalización de los equipos contratados para desarrollarla, los funcionarios que las organizan y los legisladores que deberían tomarlas en cuenta en sus decisiones, sobre todo las presupuestales. Aún no se cuenta con un avance suficiente de la profesionalización, entendida como la creación de una comunidad de expertos, que cuente con una revisión propia del conocimiento alcanzado, con base en criterios de calidad aceptados y el desarrollo de foros de discusión especializados.

En el caso mexicano, se había presentado el extenso recorrido de antecedentes que culminaba con las evaluaciones externas, las auditorías de desempeño llevadas a cabo a Progres a partir de 1998 y la generalización de las mismas a todos los programas federales que recibieran recursos fiscales y estuvieran sujetos a RO, implementada desde 2001. La presente versión incorpora los importantes acontecimientos posteriores: a nivel federal, la reglamentación de la LGDS y creación del Coneval en 2006, así como la preparación de nuevos lineamientos para guiar los diferentes tipos de evaluación y la propuesta de la SHCP para instrumentar el SED; a nivel del GDF, la correspondiente reglamentación de la LDSDF, también de 2006, y la instalación de Evalúa DF en 2008.

En consecuencia, en el presente trabajo se caracterizan las evaluaciones federales realizadas, contrastando los periodos 2001-2005 y 2006 a la fecha, para identificar avances y también desafíos pendientes. Entre los primeros, se resalta un importante mejoramiento del marco normativo, una lenta pero

progresiva profesionalización de los evaluadores, una mejor organización del sistema de evaluaciones externas y de la calidad de los trabajos, y un esfuerzo por lograr mayor congruencia en el diseño y rediseño de programas mediante la aplicación del marco lógico.

Entre los múltiples asuntos no resueltos persisten muchos identificados desde el inicio de los procesos: contar con un fondo independiente de recursos para la contratación de evaluadores al margen del programa evaluado que permita respetar su autonomía financiera e institucional para garantizar su objetividad, hacerlo siempre mediante licitación pública, avanzar en la realización de evaluaciones participativas centradas en el aprendizaje de funcionarios y ciudadanos, mejorar los esquemas de muestreo y presentarlos en los informes, emitir valoraciones fundamentadas sobre la gestión del programa, realizar evaluaciones de impacto plurianuales y de conjuntos de programas, profesionalizar el campo, aplicar metaevaluaciones, y realizar un mayor esfuerzo por institucionalizar la cultura de la evaluación.

En relación con el componente de participación social, mientras a nivel federal México sigue cuestionándose sobre su conveniencia, en diversos países constituye una estrategia ampliamente aceptada (países escandinavos, Suiza, Francia, Canadá, etcétera), siendo su preocupación actual, la forma de mejorar la calidad de la misma. Por ello se insiste en la necesidad de poner a prueba su utilidad en nuestro contexto, aplicando una estrategia incremental. Su ausencia podría considerarse un síntoma de que en realidad el gobierno no se está preocupando por impulsar el aprendizaje democrático directo, por contar con ciudadanos informados, críticos y propositivos, sino sólo por utilizar las evaluaciones para racionalizar el gasto social y ganar legitimidad.

En el caso de las evaluaciones realizadas al GDF, éstas se presentan como una práctica novedosa y emergente frente al modelo tradicional federal, basada en propuestas y procesos de planeación que implican diferencias de valores y de criterios de organización y decisión.

Resaltan entre dichas diferencias en la organización de los procesos: la preparación del proceso inicial mediante la impartición de cursos y diplomados en evaluación y la intención de elaborar guías para facilitar el trabajo; la evaluación de programas, pero también de políticas, seleccionados con base en criterios explícitos y sobre aspectos seleccionados; la realización de dos metaevaluaciones; la convocatoria por concurso abierto entre los especialistas inscritos en su

registro de evaluadores; la limitación del número de evaluaciones simultáneas a realizar por un mismo equipo; el presupuesto preestablecido y transparente en la convocatoria; las recomendaciones efectuadas por el Comité con carácter vinculatorio; y la propuesta de amplia difusión de resultados en periódicos, revistas, y de ser posible, libros, además de internet.

Asimismo, se identifican diferencias en los métodos de evaluación empleados: evaluaciones construidas en función de las características específicas del programa; reconocimiento de la existencia de múltiples y subjetivas verdades; contribución tanto de los métodos cuantitativos como de los cualitativos (interpretación, comprensión); consideración de los estudios experimentales como muy difíciles, costosos, con connotaciones éticas negativas y sin los resultados técnicos esperados; énfasis en la necesidad de contar con equipos interdisciplinarios pertinentes al programa; y realización de trabajos especializados pero con promoción de la participación social para generar aprendizaje ciudadano y credibilidad de la evaluación.

Lo anterior no significa que no se detecten también asuntos mejorables: diseño de TR; tiempo destinado al trabajo de evaluación; precisión en marcos teóricos, métodos y técnicas utilizadas; acompañamientos homogéneos; mayores recursos y montos más diversificados en función de la complejidad de la evaluación; mayor experiencia de los equipos evaluadores; contacto previo con funcionarios para facilitar acuerdos y entrega de información; mayor énfasis en la evaluación de la operación y su congruencia con el diseño; intensificación de la participación social; mayor valoración del análisis y emisión de juicios sobre la gestión; propuestas más específicas; información a los evaluadores de los procesos posteriores a la entrega de su informe final; mejoras en el sistema de seguimiento de recomendaciones; y mayor número de acciones que contribuyan a incrementar la cultura de la evaluación.

Por último, se revisa el tema de las metaevaluaciones a nivel teórico, conceptual, metodológico y empírico. En torno de los tres primeros aspectos se incorporan algunas pocas propuestas nuevas, especialmente las provenientes de Cuba y Argentina. Ante el limitado número de experiencias internacionales mencionadas, especialmente en Francia, España, Canadá y Costa Rica, en la presente edición se describen ampliamente los trabajos realizados en México en los últimos años, demostrativos de un avance sustantivo, y se analizan tres de ellos. Si consideramos que desde fines de década de 1960 Scriven ya trabajaba en

la realización de metaevaluaciones en Estados Unidos, tenemos que reconocer un avance muy lento del tema, en el que comienzan a diferenciarse enfoques complementarios de metaevaluación: metodológico, hermenéutico, político, deontológico y organizativo.

Se insiste metodológicamente en la necesidad de analizar no sólo los informes de evaluación sino también la organización y funcionamiento del sistema en su conjunto, incluyendo factores contextuales, que también contribuyen a impactar la calidad del trabajo realizado.

Como ya se mencionó, a nivel federal aún persisten muchos problemas, pero se reconoce una sensible mejoría de la calidad de las evaluaciones y una mejora en su organización y normatividad a partir de la creación del Coneval en 2006. A nivel del GDF, se percibe un esfuerzo importante por definir un modelo propio y, dado su reciente inicio de actividades, se considera que ha tenido un desempeño que, aunque mejorable, se valora como satisfactorio. También se ha mostrado un mayor interés en la realización de metaevaluaciones que contribuyan a mejorar sus procesos y la calidad de sus productos.

En síntesis, se concluye que el importante desarrollo alcanzado en los últimos años en materia de evaluación y metaevaluación en las políticas y programas públicos en México aún presenta problemas organizacionales, de calidad técnico-metodológica y de participación social.

Además, tanto el nivel federal como el del GDF muestran una urgente necesidad de contar con mayor independencia de sus respectivos gobiernos para llevar a cabo su importante función y de fortalecer la utilización de los resultados de las evaluaciones. Si esta última no se concreta se corre el peligro de hacer de la evaluación un mero rito, que incrementa el trabajo administrativo, crea un nuevo mercado de trabajo para especialistas, pero no se integra al proceso decisorio. Por dicha razón, la institucionalización de procesos de evaluación en el país no garantiza una verdadera gestión por resultados (ABR) que incorpore el conocimiento generado para un mejor diseño e instrumentación de sus políticas y programas.

Finalmente, las propuestas formuladas para coadyuvar a la resolución de los problemas identificados en el trabajo han sido planteadas con la intención de que México aproveche la enorme oportunidad que tiene actualmente de mejorar la calidad de los trabajos de evaluación e integrar sus resultados a la gestión pública, como forma de incidir más efectivamente en el alivio del más urgente problema nacional: el incremento de la pobreza registrado en los últimos años.



## ANEXO

### Normas para la evaluación de programas del Joint Committee on Standards for Educational Evaluation (segunda edición, 1994)

#### **Normas de utilidad**

Las normas de utilidad pretenden asegurar que una evaluación servirá a las necesidades de información de los usuarios pretendidos.

*U1 Identificación de los agentes críticos (stakeholders).* Se deben identificar las personas implicadas o afectadas por la evaluación, de manera que sus necesidades puedan ser cubiertas.

*U2 Credibilidad del evaluador.* Las personas que dirijan una evaluación deben ser merecedoras de confianza y competentes para realizar una evaluación, de manera que los hallazgos de la evaluación consigan la máxima credibilidad y aceptación.

*U3 Selección y alcance de la información.* La información recopilada debe ser seleccionada ampliamente para contestar preguntas pertinentes sobre el programa y debe responder a las necesidades e intereses de clientes y otros agentes críticos.

*U4 Identificación de valores.* Las perspectivas, los procedimientos y la fundamentación que se utilicen al interpretar los resultados deben describirse con cuidado, de manera que los fundamentos de los juicios de valor sean claros.

*U5 Claridad del informe.* Los informes de evaluación deben describir claramente el programa que está siendo evaluado, incluyendo su contexto, y los propósitos, procedimientos y resultados de la evaluación, de manera que se ofrezca información esencial y fácilmente comprensible.

*U6 Oportunidad y difusión del informe.* Los hallazgos intermedios significativos y los informes de evaluación se deben difundir a los usuarios deseados, de forma que puedan ser utilizados de manera oportuna.

*U7 Impacto de la evaluación.* Las evaluaciones se deben planificar, dirigir y comunicar de forma que animen a los agentes críticos a su cumplimiento, de manera que aumente la probabilidad de que se use la evaluación.

### **Normas de factibilidad/viabilidad**

Las normas de factibilidad tienen el propósito de asegurar que una evaluación sea realista, prudente, diplomática y frugal.

*F1 Procedimientos prácticos.* Los procedimientos de la evaluación deben ser prácticos, de manera que las interrupciones sean mínimas mientras se obtiene la información necesaria.

*F2 Viabilidad política.* La evaluación se debe planificar y dirigir anticipando las diferentes posturas de diversos grupos de interés, de manera que se pueda obtener su colaboración y así se puedan advertir y contrarrestar posibles intentos de restringir las operaciones de evaluación o de sesgar o abusar de los resultados por parte de estos grupos.

*F3 Coste Efectividad.* La evaluación debe ser eficiente y producir una información de valor suficiente, de manera que se justifiquen los recursos empleados.

### **Normas de corrección/legitimidad**

Son las normas dirigidas a asegurar que la evaluación se dirija legal y éticamente y con la debida atención al bienestar de aquellos implicados en la evaluación, así como a aquellos afectados por sus resultados.

*C1 Orientación de servicio.* Las evaluaciones se deben diseñar para asistir a las organizaciones a tratar y cubrir efectivamente las necesidades de todo el espectro de destinatarios.

*C2 Acuerdos formales.* Las obligaciones de las partes formales de la evaluación (qué se hará, cómo, por quién y cuándo) se deben acordar por escrito, de forma que dichas partes estén obligadas a seguir todas las condiciones del acuerdo o, por el contrario, renegociarlo formalmente.

*C3 Derechos de los sujetos.* Las evaluaciones se deben diseñar y dirigir de tal forma que se respeten y protejan los derechos y el bienestar de los sujetos.

*C4 Interacciones humanas.* Los evaluadores deben respetar la dignidad y el valor humano en sus interacciones con otras personas asociadas a la evaluación, de manera que no amenace o dañe a los participantes.

*C5 Valoración completa y justa.* La evaluación debe ser completa y justa en su análisis y en su registro de fortalezas y debilidades del programa evaluado, de tal modo que las fortalezas sean reforzadas y las áreas problemáticas tratadas.

*C6 Revelación de los resultados.* Las partes formales de la evaluación deben asegurar que el conjunto completo de resultados de la evaluación junto con sus pertinentes limitaciones son accesibles para las personas afectadas por la evaluación y para cualquier otra persona con derechos legales para recibir los resultados.

*C7 Conflicto de intereses.* Los conflictos de intereses deben manejarse abierta y honestamente de tal forma que no comprometan el proceso ni los resultados de la evaluación.

*C8 Responsabilidad fiscal.* La distribución y los gastos de recursos asignados por el evaluador deben reflejar procedimientos solventes de información y responsabilidad, en caso contrario el evaluador debe ser prudente y éticamente responsable de manera que los gastos se justifiquen y sean apropiados.

### **Normas de precisión**

Estas normas tienen el propósito de asegurar que una evaluación revele y proporcione información técnicamente apropiada sobre las características que determinan el valor o mérito del programa evaluado.

*P1 Documentación del programa.* El programa evaluado debe ser descrito y documentado de forma apropiada y clara, de tal manera que el programa se identifique claramente.

*P2 Análisis del contexto.* El contexto en el cual el programa existe se debe examinar con suficiente detalle de manera que se puedan identificar sus posibles influencias sobre dicho programa.

*P3 Descripción de propósitos y procedimientos.* Los propósitos y procedimientos de la evaluación deben ser seguidos y descritos con el suficiente detalle de manera que se puedan identificar y valorar.

*P4 Fuentes de información defendibles.* Las fuentes de información que se utilicen en una evaluación de programa se deben describir con suficiente detalle de forma que se pueda estimar la adecuación de la información.

*P5 Información válida.* Los procedimientos de recopilación de información se deben elegir y desarrollar y entonces poner en práctica de manera que se asegure que la interpretación a la que se ha llegado es válida para el uso que se le pretende dar.

*P6 Información fiable.* Los procedimientos de recopilación de información se deben elegir, desarrollar y poner en práctica de manera que se asegure que la información obtenida es suficientemente fiable para el uso que se le pretende dar.

*P7 Información sistemática.* La información recopilada, procesada y transmitida en una evaluación debe ser revisada sistemáticamente y cualquier error encontrado debe ser corregido.

*P8 Análisis de la información cuantitativa.* La información cuantitativa de una evaluación se debe analizar apropiada y sistemáticamente de manera que las preguntas de evaluación se contesten de forma efectiva.

*P9 Análisis de la información cualitativa.* La información cualitativa de una evaluación se debe analizar apropiada y sistemáticamente de manera que las preguntas de evaluación se contesten de forma efectiva.

*P10 Conclusiones justificadas.* Las conclusiones a las que se llega en una evaluación deben estar justificadas explícitamente de manera que los agentes críticos de la evaluación puedan ponderarlas.

*P11 Informe imparcial.* Los procedimientos para informar deben asegurar que no haya distorsión causada por sesgos y sentimientos personales de cualquiera de las partes de la evaluación, de manera que los informes reflejen de forma justa los resultados de la evaluación.

*P12 Metaevaluación.* La evaluación en sí misma se debe evaluar formativa y sumativamente a partir de éstas y otras normas pertinentes, de manera que se guíe de forma apropiada su realización y, en cuanto se termine, los agentes críticos puedan examinar de cerca sus fortalezas y debilidades.

## Bibliografía

### Libros y revistas

- Abate, Bernard (2000), *La nouvelle gestion publique*, Collection Systemès, L.G.D.J., París.
- Aguilar, Ma. José y Ezequiel Ander-Egg (1998), *Cómo elaborar un proyecto. Guía para diseñar proyectos sociales y culturales*, Colección Política, Servicios y Trabajo Social, Lumen/Humanitas, Buenos Aires.
- Aguilar V., Luis (compilación y estudio introductorio) (2000a), *El estudio de las políticas públicas*, Miguel Ángel Porrúa, México.
- (compilación y estudio introductorio) (2000b), *Problemas públicos y agenda de gobierno*, Miguel Ángel Porrúa, México.
- (compilación y estudio introductorio) (2000c), *La hechura de las políticas*, Miguel Ángel Porrúa, México.
- (2000d), “La reforma del Estado mexicano”, en Méndez, José Luis (comp.), *Lecturas básicas de administración y políticas públicas*, Colmex, México.
- Anderson (2000), *Public Policy Making*, Houghton Mifflin Co., Boston.
- Ballart, Xavier (1992), *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*, Ministerio para las Administraciones Públicas, Madrid.
- Banco Mundial (2001), *Attacking poverty*, World development report 2000/2001, Washington, DC.
- (2004), *La pobreza en México: una evaluación de las condiciones, las tendencias y la estrategia del gobierno*, Banco Mundial, México.
- Banco Interamericano de Desarrollo (2000), *Economic and social report 1998-1999*, IDB, Washington, DC.

- Bañón i Martínez, Rafael (comp.) (2003), *La evaluación de la acción y de las políticas públicas*, Díaz de Santos, Madrid.
- Barzelay, Michael (1992), *Breaking through bureaucracy*, University of California Press, California.
- (2001), *The New Public Management. Improving research and policy dialogue*, University of California Press, Russell Sage Foundation, Nueva York.
- Bemelmans-Videc, Marie Louise, Ray Rist y Evert Vedung (eds.) (1998), *Carrots, sticks and sermons*, Colección Comparative Policy Analysis, Ray Rist editor, Estados Unidos.
- Bichmann, Rifkin y Shrestha (1989), “Cómo medir la participación de la comunidad”, *Foro Mundial de la Salud*, vol. 10.
- Blanet, J. et al. (eds.) (2000), *Learning for change: issues and experiences in participatory monitoring and evaluation*, Intermediate Technology Publications, Londres.
- Boltvinik, Julio y E. Hernández (1999), *Pobreza y distribución del ingreso en México*, Siglo XXI Editores, México.
- Bourdieu, Pierre (dir.) (1993), *La misère du monde*, Du Seuil, Col. Libre Examen, París.
- Boyle, Richard y Donald Lemaire (eds.) (1998), *Buildong effective evaluation capacity*, Colección Comparative Policy Analysis, Ray Rist editor, Estados Unidos.
- Bresser Pereira, Luis et al. (2004), *Política y gestión pública*, FCE, Buenos Aires.
- Brewer, Garry y Peter deLeon (1983), *The foundations of policy analysis*, Dorsey Press, Estados Unidos.
- Brown, Deryck (enero 2001), “Evaluación de políticas y programas de la región del Caribe de habla inglesa: problemas conceptuales y prácticos”, *Serie Gestión Pública*, núm. 9, CEPAL/ILPES, Santiago de Chile.
- Brugué, Quim (junio 2004), “Modernizar la administración desde la izquierda: burocracia, nueva gestión pública y administración deliberativa”, *Revista del CLAD*, Reforma y Democracia, núm. 29.
- Bustelo, María (2003), “¿Qué tiene de específico la metodología de la evaluación?”, en Bañón (comp.) *La evaluación de la acción y de las políticas públicas*, Díaz de Santos, Madrid.
- (2004), *La evaluación de políticas públicas de igualdad de género en España*, Catarata, Madrid.
- Campbell, D. y J. Stanley (2005), *Diseños experimentales y cuasiexperimentales en la investigación social*, Amorrortu, Argentina.
- Cardozo Brum, Myriam (enero-junio de 1993), “La evaluación de las políticas públicas: problemas, metodologías, aportes y limitaciones”, *Revista de Administración Pública*, núm. 84, INAP, México.

- Cardozo Brum, Myriam (julio de 1999), “La evaluación en las actividades no lucrativas. Revisión de experiencias recientes en países desarrollados y situación en México”, *Revista Administración y Organizaciones*, año 1, núm. 2, UAM-Xochimilco, México.
- (enero-agosto de 2000), “Gerencia pública y participación social”, *Revista Acta Sociológica*, núms. 28-29, FCPyS-UNAM, México.
- (enero-marzo, 2005), “Procesos de evaluación y participación social en el marco de la gerencia pública necesaria para América Latina”, *Revista Venezolana de Gerencia*, núm. 10, Venezuela.
- (diciembre de 2005), “Neoliberalismo y eficiencia de los programas sociales en México”, *Revista Política y Cultura*, núm. 24, UAM-Xochimilco, México.
- (2006a), *La evaluación de políticas y programas públicos. El caso de los programas de desarrollo social en México*, Colección Conocer para Decidir, Cámara de Diputados/ Miguel Ángel Porrúa, México.
- (enero-abril de 2006b), “Políticas de lucha contra la pobreza en México. Principales resultados y limitaciones”, *Fermentum, Revista Venezolana de Sociología y Antropología*, año 16, núm. 45, ULA, Venezuela.
- (enero-abril, 2009), “La institucionalización de una cultura de la evaluación en la administración pública mexicana”, *Revista Convergencia*, año 16, núm. 49, UAEM.
- (2009), *Evaluación y metaevaluación en los programas mexicanos de desarrollo social*, Colección Teoría y análisis, UAM-Xochimilco, México.
- (enero-junio, 2011), “Organización y métodos alternativos en la evaluación de políticas y programas sociales. La experiencia consolidada y la emergente en el caso de México”, *Revista Gestión y Análisis de Políticas Públicas*, núm. 5, INAP, España.
- (2012), *Sistematización de la experiencia de Evalúa DF 2008 (metaevaluación y autoanálisis)*, UAM-Xochimilco/Evalúa DF, México.
- CEPAL (junio de 1998), *Gestión de programas sociales en América Latina*, vol I, Serie Políticas Sociales, núm. 25, Santiago de Chile.
- (2003), *Social panorama of Latin America 2002-2003*, Naciones Unidas.
- CIME-Primer Ministro-Comissariat Général du Plan (1997), *La loi quinquennale relative au travail, à l'emploi et à la formation professionnelle*, la Documentation Française, Francia.
- CNE-Comissariat Général du Plan (2002), *La politique de lutte contre le SIDA 1994-2000. Rapport de l'instance d'évaluation présidée par Christian Rollet*, la Documentation Française, Francia.
- Cohen, E. y R. Franco (1993), *Evaluación de proyectos sociales*, Siglo XXI Editores, México.

- Cohen, Ernesto (2001), “Reforma del Estado, modernización administrativa y evaluación del desempeño de la gestión”, en *Los desafíos de la reforma del Estado en los programas sociales: tres estudios de caso*, Serie Políticas Sociales, núm. 45, CEPAL, Santiago de Chile.
- Commailles, Jacques y Bruno Jobert (dirs.) (1998), *Les métamorphoses de la régulation politique*, núm. 24, Droit et Societé, Francia.
- Conde, Carola (2007), “Evaluación de programas sujetos a reglas de operación. ¿Un ejemplo de rendición de cuentas horizontal?”, en *Colección Documentos de Investigación*, El Colegio Mexiquense, México.
- Consejo Científico del CLAD (2000), *La responsabilización en la nueva gestión pública latinoamericana*, CLAD, Venezuela.
- Cooke y Kothary (2001), *Participation: the new tyranny?*, Zed Books, Londres.
- Crozier, Michel (1974), *El fenómeno burocrático*, Amorrortu, Buenos Aires.
- (1989), *Estado modesto, Estado moderno*, FCE, México.
- Crozier, M y J.C. Thoenig (1975), “La régulation des systemes organisés complexes”, *Revue Française de Sociologie*, XVI, Francia.
- Cunill, Nuria (1995) “La rearticulación de las relaciones Estado-sociedad: en búsqueda de nuevos sentidos”, *Revista del CLAD*, Reforma y Democracia, núm. 4.
- (1997), *Repensando lo público a través de la sociedad: nuevas formas de gestión pública y representación social*, CLAD, Venezuela.
- Cunill, Nuria y Sonia Ospina (2003), *Evaluación de resultados para la gestión pública moderna y democrática: experiencias latinoamericanas*, CLAD-AECI-MAP-FIIAPP. Venezuela.
- (2008), *Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina*, Proyecto CLAD/Banco Mundial.
- Chanut, Véronique (2002), “L'évaluation: affaire d'État ou question d'organisations?”, *Politiques et Management Public*, vol. 2, núm. 4, número especial, Institut de Management Public, Francia.
- Chauviré, C. y O. Fontaine (2003), *Le vocabulaire de Bourdieu*, Ellipses, Francia.
- Christensen, Tom y Per Laegreid (2001), “La nueva administración pública: el equilibrio entre la gobernanza política y la autonomía administrativa”, *Revista Gestión y Política Pública*, vol. X, núm. 1, CIDE, México.
- DeBono, E. (1999), *Six thinking hats*, Backbay, Londres.
- Deleau, M., J.P. Nioche, P. Penz y R. Poinsard (1986), *Evaluer les politiques publiques: méthodes, déontologie, organisation*, La Documentation Française, Francia.
- Díaz, Luis Fernando (2002), “La metaevaluación y su método”, *Revista de Ciencias Sociales*, año XLIII, núm. 92-93, Universidad de Costa Rica, Costa Rica.

- Díaz, Luis Fernando y Rosay Rosales (2003), *Metaevaluación: Evaluación de la evaluación de políticas, programas y proyectos sociales*, Universidad Estatal a Distancia, San José de Costa Rica.
- Duran, Patrice y E. Monnier (abril, 1992), “Le développement de l'évaluation en France. Nécessités techniques et exigences politiques”, *Revue Française de Science Politique*, vol. 40, núm. 2.
- Duran, Patrice y J.C. Thoenig (agosto de 1996), “L'État et la gestion territoriale”, *Revue Française de Science Politique*, Francia.
- Duran, Patrice (1999), *Penser l'action publique*, Droit et Société, vol. 27, Francia.
- (diciembre de 2002), “L'évaluation, au risque de son institutionalisation”, *Politiques et Management Public*, vol. 2, núm. 4, número especial, Institut de Management Public, Francia.
- Ernst & Young (2008), *Barómetro 2008 de la evaluación de políticas públicas en Europa*, Francia.
- Ferlie, E. et al. (1996), *The new public management in action*, Oxford University Press, Oxford.
- Fernández Toro, Julio César (junio de 2004), “Presentación”, en *NotiClad*, núm. 23, CLAD, Venezuela.
- Fink, A. y J. Kosecoff (1978), *An evaluation primer*, Washington Capitol Publications, Washington.
- Fisher, Frank (1997), *Evaluating public policy*, Nelson Hall Publishers, Chicago.
- Frade Rubio, Laura (comp./coord.) (2000), *¿Equidad, participación y correspondencia?*, Campaña El Banco Mundial en la Mira de las Mujeres, México.
- Freyssinet, Jacques (1998), “La fonction sociale de l'évaluation des dispositifs d'insertion professionnel des jeunes. Une comparaison européenne”, en Kesler, Marie Christine et al., *Evaluation des politiques publiques*, LHarmattan, París.
- García Sánchez, Ester (2009), “Metaevaluación de políticas públicas: una visión desde la ciencia política”, *Revista Reforma y Democracia*, núm. 43, CLAD, Venezuela.
- Garde, Juan Antonio (2004), “La evaluación de políticas y su institucionalización en España”, *Revista Gestión y Análisis de Políticas Públicas*, núms. 30-31, INAP, España.
- Glass, Gene (1976), “Primary, secondary and meta-analysis of research”, *Educational Researcher*, núm. 5, pp. 3-8.
- Goldsworthy, David (1998), “Le contrôle du bon usage des fonds publics par le National Audit Office”, en Conseil Scientifique de l'Évaluation, *L'évaluation en développement 1997*, La Documentation Française, Francia.
- Guba, E.G. (1987), *Naturalistic evaluation. New direction for program evaluation*, núm. 34, Jossey Bass, San Francisco.

- Guba, E.G. (1990), *The paradigm Dialog*, Newsbury Park Publications.
- y V. Caracelli (1989), *Fourth generation evaluation*, Sage, Estados Unidos.
- Guba, E. y Lincoln (1981), *Effective evaluation*, Jossey Bass, San Francisco.
- Guerrero, Juan Pablo (primer semestre de 1995), “La evaluación de políticas públicas: enfoques teóricos y realidades de nueve países desarrollados”, *Revista de Gestión y Política Pública*, núm. 1, vol. IV, CIDE, México.
- Guerrero, Omar (1998), “El Management Público: una torre de Babel”, *Revista Convergencia*, año 5, núm. 17, UAEM, México.
- (1999), *Del Estado gerencial al Estado cívico*, UAEM/Miguel Ángel Porrúa, México.
- (2003), *Gerencia pública en la globalización*, UAEM/Miguel Ángel Porrúa, México.
- (enero-marzo de 2004), “El mito del nuevo ‘management’ público”, *Revista Venezolana de Gerencia*, año 9, núm. 25, Universidad del Zulia, Venezuela.
- Guimaraes, Roberto (2008), “Estado, mercado y democracia: oportunidades y límites de la participación en el fortalecimiento de la gobernabilidad democrática”, *Revista del CLAD Reforma y Democracia*, núm. 40, CLAD, Venezuela.
- Hespel, Véronique (diciembre de 2002), “L’avenir institutionnel de l’évaluation des politiques publiques”, *Politiques et Management Public*, vol. 2, núm. 4, número especial, Institut de Management Public, Francia.
- Hofstède, G. (1980), *Management control of public and not for profit activities*, International Institute for Applied Systems Analysis, Austria.
- Hood, C. (1994), *Explaining economic policy reversals*, Open University Press, Buckingham.
- Hood, C. y M. Jackson (1997), *La argumentación administrativa*, FCE, México.
- Isaac, W. (1999), *Dialogue and the art of thinking together: a pioneering approach to communicating in bussines and in life*, Doubleday, Nueva York.
- Joint Committee on Standards for Educational Evaluation (1994), *The program evaluation standards*, Sage Publications, Thousand Oaks.
- Jobert, Bruno (1998), “La régulation politique: le point de vue d’un politiste”, en Commailles, J. y B. Jobert (eds.), *Les métamorphoses de la régulation politique*, Librairie Général de Droit et de Jurisprudence, París.
- Kessler, Marie-Christine et al. (1998), *Evaluation des politiques publiques*, L’Harmattan, París.
- Kliksberg, Bernardo (1989), “¿Cómo formar gerentes sociales? Elementos para el diseño de estrategias”, en B. Kliksberg (comp.), *¿Cómo enfrentar la pobreza?*, Grupo Editor Latinoamericano, Argentina.

- Kliksberg, Bernardo (1994), *Pobreza. El drama cotidiano. Clave para una nueva gerencia social eficiente*, CLAD/PNUD/Tesis-Norma, Venezuela.
- Lacasse, M. François (1993), “Gestion et émasculon des évaluations: le cas de Canada”, *Revue Française d’Administration Publique*, núm. 66, Francia.
- Lakatos, Imra (1989), *La metodología de los programas de investigación científica*, Alianza Editorial, Madrid.
- Lascoumes, P y M. Setbon (1996), *L’évaluation pluralista des politiques publiques. Enjeux)), pratiques, produits*, GAPP/CNRS/CGP, Francia.
- Lascoumes, Pierre (1998), “Pratiques et modèles de l’évaluation” y “Usages et retombées de l’évaluation”, en Kesler, Marie Christine *et al.*, *Evaluation des politiques publiques*, L’Harmattan, París.
- Lasswell, Harold (1951), “La orientación hacia las políticas”, en Aguilar V., Luis (2000a).
- (1971), “La concepción emergente de las ciencias políticas”, en Aguilar V., Luis, (2000a),
- Leeuw, Frans, Ray Rist y Richard Sonnichsen (eds.) (2000), *Can government learn?*, Colección Comparative Policy Analysis, Ray Rist editor, Estados Unidos.
- Lindblom, Charles (1959), “The science of muddling through”, *Public Administration Review*, vol. 19, Estados Unidos.
- Lindblom y Wood-House (1993), *The policy making process*, Prentice Hall, Nueva York.
- Long, Norman (2007), *Sociología del desarrollo: una perspectiva centrada en el actor*, CIESAS/Colegio de San Luis, México.
- Llerenas, Vidal (2007), “Contraloría social: ejerciendo el derecho a la rendición de cuentas en los programas de desarrollo social”, *Revista del CLAD Reforma y Democracia*, núm. 39, CLAD, Venezuela.
- Majone, Giandomenico (1996), “Política pública y administración: ideas, intereses e instituciones”, en Oxford University, *A new handbook of political science*, Oxford University Press, Oxford.
- Martin, Gérard y Amédine Ruffiot (junio de 2001), “La commande d’évaluation de politiques sociales territoriales, entre mythes et apprentissages”, *Revue Politiques et Management Public*, vol. 19, núm. 2, Francia.
- Mejía Lira, José (2003), *La evaluación de la gestión y las políticas públicas*, Miguel Ángel Porrúa, México.
- Mény, Yves y J.C. Thoenig (1995), *Las políticas públicas*, Ariel, Barcelona.
- Monnier, Eric (1989), “Évaluation des politiques publiques: méthodes et pratiques”, *Problèmes politiques et sociaux*, núm. 599.

- Monnier, Eric (1991), “Objectifs et destinataires des évaluations”, *Documentación Administrativa*, núms. 224-225, España.
- (1995), *Evaluación de la acción de los poderes públicos*, Instituto de Estudios Fiscales, España.
- Monnier, Eric y Vincent Spenlehour (septiembre, 1992), “L'évaluation dans le triangle de la décision. Opinions singulières et processus pluraliste”, *Politiques et management publique*, vol. 10, núm. 3.
- Moreno, Salvador (2007), “Los insumos para la evaluación de las políticas públicas: la calidad de la información”, *Reporte CESOP*, núm. 9, LX Legislatura, Cámara de Diputados, México.
- Mosqueira, Edgardo (2007), “De la evaluación por productos a la evaluación por resultados e impactos: análisis de las experiencias recogidas del proyecto de modernización del Estado argentino”, *Revista del CLAD Reforma y Democracia* núm. 38, CLAD, Venezuela.
- Muller, Pierre (1994), *Les politiques publiques*, Collection Que sais-je?, Presses Universitaires de France, París.
- Murciano, J. (2007), “Evaluar la evaluación: una herramienta para la mejora. Aplicación al caso de la Iniciativa Equal en Andalucía”, en M.J. Terol Becerra (dir.), *Nuevas políticas públicas. Anuario multidisciplinar para la modernización de las administraciones públicas*, núm. 3, Instituto Andaluz de Administración Pública, Sevilla.
- Naciones Unidas (2001), *World public sector report: globalization an the state*.
- (2002), *Informe de Desarrollo Humano*.
- Nioche, Jean Paul (febrero de 1982), “De l'évaluation à l'analyse des politiques publiques”, *Revue Française de Science Politique*, núm. 1, Francia pp. 32-61.
- (1985), “Les politiques publiques”, *Traité de Science Politique*, tomo 4, PUF, París.
- (abril-junio de 1993), “L'évaluation de politiques publiques en France: ‘fast food’, recettes du terroir ou cuisine internationale?”, *Revue Française d'Administration Publique*, núm. 66, Francia.
- Nioche, J.P. y R. Poincard (1994), *L'évaluation des politiques publiques*, Economica, París.
- Nonaka, I. y T. Nishiguchi (2001), *Knowledge emergence: social, technical and evolutionary dimensions of knowledge creation*, Oxford University Press, Nueva York.
- OCDE (1995), *Governance in transition. Public management reforms in OECD countries*.
- (2001), *Evaluation feedback for effective learning and accountability*.
- Ochoa, Sara (2007), “Perspectiva sobre la evaluación y gestión pública”, *Reporte CESOP*, núm. 9, LX Legislatura, Cámara de Diputados, México.

- O'Donnell, Guillermo y O. Oszlak (1976), *Estado y políticas estatales en América Latina: hacia una estrategia de investigación*, CEDES, Argentina.
- Osborne y Gabler (1994), *La reinención del gobierno*, Paidós, Buenos Aires.
- Osborne y Plastrik (1998), *La reducción de la burocracia. Cinco estrategias para reinventar el gobierno*, Paidós, España.
- Ospina, Sonia (2001), “Evaluación de la gestión pública: conceptos y aplicaciones en el caso latinoamericano”, *Revista Reforma y Democracia*, núm. 19, CLAD, Venezuela.
- Osuna, J.L. et al. (2000), *Guía para la evaluación de políticas públicas*, Instituto de Desarrollo Regional, Sevilla, España.
- Osuna, J.L. y C. Vélez (2001), “Evaluation as an instrument for administrative reform in Spain”, *Revista Internacional de Estudios Políticos*, volumen Especial, NUSEG/UERJ, Río de Janeiro, Brasil.
- Osuna, J.L., C. Vélez y J. Murciano (2006), “Programación y evaluación pública: un triángulo complejo”, *Ekonomiaz*, Revista Vasca de Economía, núm. 60, Servicio Central de Publicaciones del Gobierno Vasco, Viztoria-Gasteiz, España, pp. 76-97.
- Oszlak, Oscar (1980), *Políticas públicas y regímenes políticos: reflexiones a partir de algunas experiencias latinoamericanas*, Estudio CEDES, vol. 3, núm. 2, Argentina.
- (2003), “Profesionalización de la función pública en el marco de la Nueva Gestión Pública”, en Arellano, David, Rodrigo Egaña, Oscar Oszlak y Regina Pacheco, *Retos de la profesionalización de la función pública*, CLAD/AECI/MAP/FIAPP, Venezuela.
- Parlett, M. y D. Hamilton, (1985), “La evaluación como iluminación”, en Gimeno Sacristán y A. Pérez Gómez (comps.), *La enseñanza: su teoría y su práctica*, Akal, Madrid.
- Patton, M.Q. (1997), *Utilization-focused evaluation. The new century text*, Sage Publications, Thousand Oaks.
- Perret, Bernard (1993), “Refléxions sur les différents modèles d'évaluation”, *Revue Française d'Administration Publique*, núm. 66, Francia.
- Pollit, Christopher (1993), *Managerialism and the public services*, Blakwell, Oxford.
- (1998), “Institutions et usages de l'évaluation au Royaume Uni: une vue d'ensemble”, en Conseil Scientifique de l'Évaluation, *L'évaluation en développement 1997*, La Documentation Française, Francia.
- Quade, E. (1975), *Analysis for public decisions*, American Elsevier, Estados Unidos.
- Rieper y Toulemonde (1996), *Politics and practices of inter-governmental evaluation*, Comparative Policy Analysis, Series Editor.
- Rist, Ray (1990), *Program evaluation and the management of government: patterns and prospects across eight nations*, Transaction Publishers, New Brunswick.

- Roggero, Pascal (2005), *De la complexité des politiques locales*, L'Harmattan, Francia.
- (2006), *Sociologie des représentations du pouvoir local*, Serie monográfica en Ciencias Sociales, núm. 10, Université Laurentienne, Canadá.
- Rossi y Freeman (1993), *Evaluation: a systematic approach*, Sage Publications.
- Saint-Clair, W. (abril-junio de 1993), “L'évaluation des politiques publiques au Royaume Uni”, *Revue Française d'Administration Publique*, núm. 66, Francia.
- Sanín Angel, Héctor (agosto de 1999), “Control de gestión y evaluación de resultados en la gerencia pública”, *Serie Manuales*, núm. 3, CEPAL, Santiago de Chile.
- Schwandt, Thomas y Edward Halpern (1988), *Linking auditing and metaevaluation*, Sage Publications, Newbury Park.
- Schwentesius, Rita et al. (julio-diciembre de 2007), “Metaevaluación de tres evaluaciones oficiales de la Sagarpa del programa de pagos directos (Procampo) a la agricultura mexicana”, *Revista Estudios Sociales*, año/vol. 15, núm. 030, Universidad de Sonora, Hermosillo, Sonora, México
- Scriven, Michael (1967), “The methodology of evaluation”, en Tyler et al., *Perspective of curriculum evaluation*, Rand McNally, Chicago.
- (1969), “An introduction to meta-evaluation”, *Educational Product Report*.
- (1973), “Goal free evaluation”, en House, E. (ed.), *The politics and process*, McCutchan, Nueva York.
- (1976), “Evaluation bias and its control”, *Evaluation Studies Review Annual*, vol. 1.
- (1991), *Evaluation Thesaurus*, Sage Publications, Newbury Park.
- Setbon, Michel (1998), “Introduction”, en Kesller, Marie Christine et al., *Evaluation des politiques publiques*, L'Harmattan, París.
- Simon, Herbert (1972), *El comportamiento administrativo*, Aguilar, España.
- (1957), *Models of man, social and rational*, Wiley, Nueva York.
- Skoufias, Emmanuel (2001), “Progresas and its impacts on the human capital and welfare of households in rural Mexico; A Synthesis of the results of an evaluation by IFPRI”, International Food Policy Research Institute, Washington, DC.
- Skoufias, Emmanuel y Susan Parker (2001), “Conditional cash transfers and the impact of child work and schooling: evidence from the Progresas program in Mexico”, *Economía: Journal of Latin America and the Caribbean Economic Association*, vol. 2, núm. 1.
- Solarte Pazos, Leonardo (2004), *La evaluación de políticas públicas en el Estado liberal*, Universidad del Valle, Cali, Colombia,
- Stake, R.E. (2006), *Evaluación comprensiva y evaluación basada en estándares*, Graó, Barcelona.

- (2010), “El desarrollo de centros de evaluación y la formación de evaluadores”, *Gestión y Análisis de Políticas Públicas*, nueva época, trabajo monográfico sobre evaluación educativa coordinado por S. Kushner y E. García Sánchez, núm. 3.
- Stame, Nicoletta (1998), “L'évaluation en Italie”, en Conseil Scientifique de l'Évaluation, *L'évaluation en développement 1997*, La Documentation Française, Francia.
- Steinbruner, John (1974), *The cybernetic theory of decision*, Princeton University Press, Nueva Jersey.
- Stufflebean, Daniel y Anthony Shinkfield (1987), *Evaluación sistemática*, Paidós, España.
- Subirats, Joan (1989), *Análisis de políticas públicas y eficacia de la administración*, INAP, España.
- Sulbrandt, José (1993), “La evaluación de los programas sociales: una perspectiva crítica de los modelos actuales”, en Kliksberg, Bernardo (comp.), *Pobreza. Un tema impostergable. Nuevas propuestas a nivel mundial*, CLAD/FCE/Programa de Naciones Unidas, México.
- Taylor, R. (1967), “Changing concepts of educational evaluation”, en R.E. Stake (comp.), *Perspectives of curriculum evaluation*, vol. 1, Rand Mc Nally, Nueva York.
- Tarrés, María Luisa (2001), *Observar, escuchar y comprender. Sobre la tradición cualitativa de la investigación social*, Colmex/Flacso/Miguel Ángel Porrúa, México.
- Thoenig, Jean-Claude (1994), “La gestion systémique de la sécurité publique”, *Revue Française de Sociologie*, XXXV, France.
- (1998), “Que peut-on apprendre s'agissant de la France de la rencontre franco-britannique?”, en Conseil Scientifique de l'Évaluation, *L'évaluation en développement 1997*, La Documentation Française, Francia.
- (1998b), “L'évaluation de l'action publique locale”, en Kesler, Marie Christine et al., *Evaluation des politiques publiques*, L'Harmattan, París.
- (segundo semestre de 2002), “La evaluación en actos. Lecciones del pasado y preguntas para el futuro”, revista *Gestión y Política Pública*, vol. XI, núm. 2, CIDE, México; y (diciembre de 2002), versión en francés, *Politiques et Management Public*, vol. 2, núm. 4, número especial, Institut de Management Public, Francia.
- UAM-Xochimilco (otoño de 1996), “Políticas públicas y sociedad”, revista *Política y Cultura*, núm. 7, México.
- Uvalle, Ricardo (1994), “Prólogo”, en Plasencia, Adriana, *Gerencia pública. Bases teórico-conceptuales*, Colegio Nacional de Ciencias Políticas y Administración Públicas, México.
- (1997), “Estudio introductorio”, en Hood, C y M. Jackson, *La argumentación administrativa*, FCE, México.

- (enero-marzo de 1998), “El carácter multifacético de la gestión pública contemporánea”, *Revista del Instituto de Administración Pública del Estado de México*, núm. 37, INAPEM, Toluca, México.
- (junio de 2001), “El fortalecimiento de la vida democrática: la nueva relación de la sociedad y el Estado”, *Revista Reforma y Democracia*, núm. 20, CLAD, Caracas, Venezuela.
- Vargas, Germán (2007), *Metaevaluación de los programas de la Secretaría de Agricultura (Sagarpa), dirigida a productos agrícola básicos. Resultados generales*, CEDRSSA, México.
- Vedung, E. (1997), *Evaluación de políticas públicas y programas*, Ministerio de Trabajo y Asuntos Sociales, Instituto Nacional de Servicios Sociales, Madrid.
- Weisbrod (1978), en Layard, Richard (comp.), *Análisis costo-beneficio*, FCE, México.
- Weiss, Carol (1982), *Investigación evaluativa*, Trillas, México.
- (1998), *Evaluation. Methods for studying programs and policies*, Prentice Hall, Estados Unidos.
- (1999), “Las investigaciones de políticas: ¿datos, ideas y argumentos?”, en Wagner, Peter, Carol Weiss, Bjorn Wittrock y Hellmut Wollman (comps.), *Ciencias sociales y Estados modernos. Experiencias nacionales e incidencias teóricas*, FCE, México.
- Wilson, Woodrow (junio de 1887), “The study of administration”, *Political Science Quarterly*, vol. II, núm. 2.
- Zaltsman, Ariel (junio, 2004), “La evaluación de resultados en el sector público argentino: un análisis a la luz de otras experiencias en América Latina”, *Revista del CLAD, Reforma y Democracia*, núm. 29.
- Zapico, Eduardo (2005), “Desarrollo integrado de la evaluación y el presupuesto por resultados: un camino largo y pedregoso con destino incierto”, en Koldo Echebarría et al., *Responsabilización y evaluación de la gestión pública*, CLAD/AECI/MAP/FIAPP, Venezuela.
- Ziccardi, Alicia y Homero Saltalamachia (1997), *Metodología de evaluación del desempeño de los gobiernos locales en ciudades mexicanas*, Instituto de Investigaciones Sociales, UNAM, México.

### **Documentos, ponencias, periódicos, tesis y consultas a internet**

- Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios (2008), “Primer año de la Agencia de Evaluación y Calidad” [[www.aeval.es/es/difusion\\_y\\_comunicacion/actualidad/actividad\\_institucional/2008-02-05.html](http://www.aeval.es/es/difusion_y_comunicacion/actualidad/actividad_institucional/2008-02-05.html)], fecha de consulta: 15 de julio de 2009.

- , Informes Generales de Actividad 2007, 2008, 2009 y 2010 [www.aeval.es], fecha de consulta: 29 de julio de 2011.
- Arellano, David y José Ramírez, “Presupuesto dirigido a resultados: los dilemas de la evaluación por desempeño en contextos latinoamericanos. Lecciones desde México” [www.clad.org.ve], fecha de consulta: 26 de julio de 2003.
- Auditoría Superior de la Federación, “Informe de resultados sobre la revisión de la cuenta de la hacienda pública federal 2000” [www.asf.gob.mx], fecha de consulta: 13 de septiembre de 2004.
- Auditoría Superior de la Federación, “Informe del resultado de la revisión y fiscalización superior de la cuenta pública 2001” [www.asf.gob.mx], fecha de consulta: 13 de septiembre de 2004.
- Auditoría Superior de la Federación, “Informe del resultado de la revisión y fiscalización superior de la cuenta pública 2002” [www.asf.gob.mx], fecha de consulta: 13 de septiembre de 2004.
- Baker, Judy (2000), “Evaluación del impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales”, Banco Mundial [www.worldbank.org], fecha de consulta: 10 de mayo del 2003.
- Banco Mundial (2004b), “México presenta avances y enfrenta retos en su lucha contra la pobreza”, nota publicada el 28 de julio [www.worldbank.org], fecha de consulta: 29 de julio de 2004.
- BID (1997), “Evaluación: una herramienta de gestión para mejorar el desempeño de los proyectos”, Washington Oficina de Evaluación del BID.
- Boltvinik, Julio (2005), “Sedeso boicotea la Ley de Desarrollo Social”, *La Jornada*, 26 de agosto.
- (2005), “Evaluación, dos modelos”, *La Jornada*, 2 de septiembre.
- (2005), “¿Controversia constitucional o acuerdo?”, *La Jornada*, 30 de septiembre.
- (2005), “Acuerdos entre Sedeso y el Legislativo”, *La Jornada*, 7 de octubre.
- (2005), “Reglamento de la Ley de Desarrollo Social”, *La Jornada*, 14 de octubre.
- (2006), “Sedeso: crisis e ilegalidad”, *La Jornada*, 20 de enero.
- (2006), “El reto de la evaluación en México”, *La Jornada*, 10 de febrero.
- (2006), “Elecciones excluyentes y controversia constitucional”, *La Jornada*, 17 de febrero.
- (2006), “Imposición del modelo social único”, *La Jornada*, 24 de febrero.
- (2006), “Fortalecer lo social frente a lo económico”, *La Jornada*, 3 de marzo.
- (2006), “La Sedeso quiere medir la pobreza”, *La Jornada*, 10 de marzo.

- Bustelo, María (2001), “La evaluación de las políticas públicas de igualdad de género de los gobiernos central y autonómicos en España: 1995-1999”, tesis doctoral, Universidad Complutense, Madrid.
- (2002), “Metaevaluation as a tool for the empowerment and development of the evaluation function in Public Administration”, ponencia presentada en la *Quinta Conferencia Europea de Evaluación*, Sevilla [[www.europeanevaluation.org](http://www.europeanevaluation.org)].
- Cardozo Brum, Myriam (2010), “Métodos cuantitativos y cualitativos en la evaluación de programas sociales”, en *Las ciencias sociales y las humanidades*, UAM-Xochimilco, México.
- (junio de 2010), “Sistematización de la experiencia de evaluación de Evalúa DF: metaevaluación y autoanálisis” [[www.evalua.df.gob.mx/transparencia/2010/met\\_inf\\_fin.pdf](http://www.evalua.df.gob.mx/transparencia/2010/met_inf_fin.pdf)], fecha de consulta: 26 de mayo de 2011.
- CCEDRSA (2 de abril de 2008), “El papel del Congreso en la evaluación de las políticas públicas rurales”, documento presentado en Mesa Redonda del mismo nombre, Cámara de Diputados, México.
- Centeno, Helen (2002), “Metaevaluación del Programa antidrogas y salud sexual y reproductiva para adolescentes, Jamaica 1999-2002: lecciones aprendidas y contribución metodológica”, tesis del máster profesional en población y salud, Universidad de Costa Rica, Costa Rica.
- Centre for European Evaluation Expertise (2002) [[www.eureval\\_c3e.fr](http://www.eureval_c3e.fr)], fecha de consulta: 4 de noviembre.
- CERTU (2002), “Évaluation des politiques publiques: faut-il quantifier pour évaluer?”, Cuaderno núm. 6, Francia [[www1.certu.fr](http://www1.certu.fr)].
- CLAD (2010), “Gestión pública iberoamericana para el siglo XXI”, XL Reunión Ordinaria del Consejo Directivo del CLAD, Santo Domingo, República Dominicana [[www.clad.org](http://www.clad.org)], fecha de consulta: 23 de febrero de 2011.
- Cohen, Ernesto (octubre de 2004), “Evaluación y monitoreo de programas sociales”, Seminario, núm. 22, Sedesol.
- Consejo Científico del CLAD (octubre de 1998), “Una nueva gestión pública para América Latina” [[www.clad.org.ve](http://www.clad.org.ve)], fecha de consulta: 4 de octubre de 2004.
- Constitución Política de los Estados Unidos Mexicanos [[www.cddhuc.gob.mx](http://www.cddhuc.gob.mx)], fecha de consulta: 22 de marzo de 2004.
- Cunill, Nuria y Sonia Ospina (2008), “Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina. Informe comparativo de 12 países”, Banco Mundial/CLAD.
- Dopico, Ileana (s/f), “Metaevaluación: ¿Por qué y para qué?” [[http://www.dict.uh.cu/Revistas/Educ\\_Sup/032003/Art030303.pdf](http://www.dict.uh.cu/Revistas/Educ_Sup/032003/Art030303.pdf)], fecha de consulta: 15 de julio de 2008.

- EES (octubre 2002), “Tres movimientos en la evaluación contemporánea: aprendizaje, teoría y pruebas, Convocatoria a la V Conferencia Bienal, España [www.europeanevaluation.org].
- Foresti, Marta et al. (2002), “The role of civil society in evaluating poverty reduction strategies and producing evidence for policy development: myth or reality?”, ponencia presentada en la V Conferencia Europea de Evaluación, Sevilla [www.europeanevaluation.org].
- Franco, Rolando (28-31 de octubre de 2003), “Descentralización, participación y competencia en la gestión social”, ponencia presentada en el *VIII Congreso Internacional del CLAD*, Panamá.
- García Sánchez, Ester (2003), “La evaluación de programas de reforma educativa en España. Tres estudios de caso desde un enfoque de metaevaluación”, tesis doctoral, Universidad Complutense de Madrid.
- (2004), “La evaluación educativa en España: balance y metaevaluación”, ponencia presentada al IX Congreso Internacional del CLAD, Madrid, 2 al 5 de noviembre.
- Guinart i Solá, Josep Maria (28-31 de octubre de 2003), “Indicadores de gestión para las entidades públicas”, ponencia presentada en el *VIII Congreso Internacional del CLAD*, Panamá.
- Hernández Licona, Gonzalo (mayo de 2003), “Sistema de evaluación y monitoreo”, trabajo presentado en la *Conferencia internacional mejores prácticas de política social*, Sedesol/Banco Mundial/BID/CEPAL, México.
- (2005), “The political side of social program evaluation”, trabajo inédito, Sedesol, México.
- Hintze, Jorge (28-31 de octubre de 2003), “Responsabilización: instrumentos de rendición de cuentas y evaluación de gestión del valor público”, ponencia presentada al *VIII Congreso Internacional del CLAD*, Panamá, Panamá.
- IDEAS [www.ideas-int.org], fecha de consulta: 23 de junio de 2011.
- IFPRI, “Mexico: evaluation of Progresá, 1997-1999” [www.ifpri.org], fecha de consulta: 15 de mayo de 2003.
- INEGI (2000), ((Niveles de bienestar en México)) [www.inegi.gob.mx], fecha de consulta: 23 de octubre de 2002.
- Jaso, Sofía y Mauricio Merino (diciembre de 1994), “Solidaridad: los matices de una política”, documento preparado para el Comité Técnico de Evaluación del Programa Nacional de Solidaridad, mimeo.
- Lajoudie, B. y J. Toulemonde (junio de 2002), “Méta-évaluation, bilan des évaluations réalisées en Limousin au cours de la période 1994-1998”, ponencia presentada a las *Cuartas Jornadas de la Société Française d'Évaluation*, Lille, Francia.

- Larrú, José María (2003), “Análisis de los resultados y metaevaluación del sistema de evaluaciones de EuropeAid”, CECOD, España [[www.cecod.net](http://www.cecod.net)], fecha de consulta: 22 de junio de 2004.
- Ley de Adquisiciones, Arrendamientos y Servicios del Sector Público, *Diario Oficial de la Federación*, 4 de enero de 2000.
- Ley de Desarrollo Social para el Distrito Federal, *Gaceta Oficial del Distrito Federal*, 11 de febrero de 2009 [<http://www.evalua.df.gob.mx>], fecha de consulta: 14 de diciembre de 2009.
- Ley de Fiscalización Superior de la Federación [[www.asf.gob.mx](http://www.asf.gob.mx)], fecha de consulta: 2 de mayo de 2003.
- Ley de Planeación del Desarrollo del Distrito Federal, *Gaceta Oficial del Distrito Federal*, 27 de enero del 2000 [<http://www.evalua.df.gob.mx>], fecha de consulta: 14 de diciembre de 2009.
- Ley Federal de Acceso a la Información Pública Gubernamental [[www.secodam.gob.mx](http://www.secodam.gob.mx)], fecha de consulta: 2 de mayo de 2003.
- Ley Federal de Fomento a las Actividades Realizadas por las Organizaciones de la Sociedad Civil [[www.cddhcu.gob.mx](http://www.cddhcu.gob.mx)], fecha de consulta: 2 de mayo de 2003.
- Ley Federal de Responsabilidades de los Servidores Públicos [[www.diputados.gob.mx/LeyesBiblio/pdf/240.pdf](http://www.diputados.gob.mx/LeyesBiblio/pdf/240.pdf)], fecha de consulta: 14 de diciembre de 2009.
- Ley General de Desarrollo Social [[www.sedesol.gob.mx](http://www.sedesol.gob.mx)], fecha de consulta: 31 de marzo de 2004.
- Ley Orgánica de la Administración Pública del Distrito Federal [[www.evalua.df.gob.mx](http://www.evalua.df.gob.mx)], fecha de consulta: 14 de diciembre de 2009.
- Longo, Francisco y Koldo Echevarría (2001), “La nueva gestión pública en la reforma del núcleo estratégico del gobierno. Experiencias latinoamericanas”, en CLAD, Serie Estudios de Casos, *Nueva gestión pública y regulación en América Latina. Balances y desafíos* [[www.clad.org.ve](http://www.clad.org.ve)], fecha de consulta: 4 de octubre de 2004.
- Marabotto, María Irma (2005), “Metaevaluación: una mirada crítica sobre la evaluación institucional”, presentación efectuada en el *Cuarto Congreso Internacional de Educación Santillana*, La calidad como meta. La evaluación como camino, Fundec, Buenos Aires.
- Monnier, Eric y Pitarelli (2000), “Benchmarking; the missing link between evaluation and management”, ponencia presentada en la Cuarta Conferencia Europea de Evaluación, Lausanne [[www.europeanevaluation.org](http://www.europeanevaluation.org)].
- Nioche, Jean Paul (junio de 1992), Conferencia sobre evaluación, impartida en CIDE, México.
- Nirenberg, Olga (28-31 de octubre de 2003), “El rol del Estado para la participación social en la evaluación”, ponencia presentada en el *VIII Congreso Internacional del CLAD*, Panamá.

- Ohana, Paul (4 de julio de 2008), “Où va l'évaluation en France?”, conferencia presentada en el *Colloque Européen L'évaluation des Politiques Publiques en Europe. Cultures et Futurs*, Estrasburgo, Francia.
- Ohana, Paul (7 de julio de 2008), “Le NAO à la française, renvoyé à plus tard?” [[http://paulohana.typedad.com/le\\_blog\\_de\\_paul\\_ohana/2008/07/le-nao-la-frana.html](http://paulohana.typedad.com/le_blog_de_paul_ohana/2008/07/le-nao-la-frana.html)], fecha de consulta: 15 de julio de 2008.
- (22 de julio de 2008), “L'évaluation des politiques publiques fait son entrée dans la Constitution” [[http://paulohana.typedad.com/le\\_blog\\_de\\_paul\\_ohana/2008/07/le-nao-la-frana.html](http://paulohana.typedad.com/le_blog_de_paul_ohana/2008/07/le-nao-la-frana.html)], fecha de consulta: 30 de julio de 2008.
- ONU (marzo de 1994), “Informe de la undécima reunión de expertos sobre el Programa de las Naciones Unidas en materia de administración y finanzas”, ONU, Nueva York.
- Osuna, José Luis y otros (2011), “Metaevaluación de las políticas evaluadas por Evalúa DF durante 2009” [[www.evalua.df.gob.mx/metaevaluaciones.html](http://www.evalua.df.gob.mx/metaevaluaciones.html)], fecha de consulta: 13 de junio de 2011.
- Osuna, J.L. (2005), “Evaluación de políticas públicas a través del Estado, la democracia y el desarrollo”, *III Congreso argentino de administración pública. Sociedad, Estado y democracia*, Tucumán, Argentina.
- PNUD (2004), “Democracia en América Latina: hacia una democracia de ciudadanos y ciudadanas” [[www.undp.org](http://www.undp.org)], fecha de consulta: 10 de octubre de 2004.
- Presupuesto de Egresos de la Federación para el Ejercicio Fiscal de 2001, *Diario Oficial de la Federación*, Primera Sección, 31 de diciembre de 2000.
- Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLAC) [[www.relacweb.com](http://www.relacweb.com)], fecha de consulta: 28 de junio de 2011.
- Red Latino Americana y del Caribe de Monitoreo y Evaluación (Redlacme) [[redlacme.org](http://redlacme.org)], fecha de consulta: 28 de junio de 2011.
- Repetto, Fabián (28-31 de octubre de 2003), “Capacidad estatal: requisito necesario para una mejor política social en América Latina”, ponencia presentada en el *VIII Congreso Internacional del CLAD*, Panamá.
- Reyna, José Luis y Alberto García Rocha (coords.) (s/f), “Metodología para la evaluación del impacto socioeconómico del Programa Nacional de Solidaridad”, Colmex.
- Rosenbaum, Allan (2002), “Program evaluation and inequality: on the need for social equity impact analysis”, *VII Congreso Internacional del CLAD sobre la Reforma del Estado y la Administración Pública*, Lisboa, 8-11 de octubre.
- Sedesol [[www.sedesol.gob.mx](http://www.sedesol.gob.mx)], fecha de consulta: 2 de mayo de 2009.
- (18 de enero de 2006), “Reglamento de la Ley General de Desarrollo Social”, *Diario Oficial de la Federación*.

- (17 de enero de 2008), “Evaluación del proceso de evaluación del programa Oportunidades”, Cuernavaca, México [<http://siteresources.worldbank.org>].
- SFP (julio de 2003), “Guía para realizar evaluaciones externas según las disposiciones emitidas por al Secodam y la SHCP y publicadas en el *Diario Oficial de la Federación*, el 30 de abril de 2002”, mimeo.
- SHCP/Secodam (30 de abril de 2002), “Acuerdo por el que se expiden los requisitos mínimos que las dependencias y entidades de la Administración Pública Federal deberán considerar para seleccionar a las instituciones académicas y de investigación u organismos especializados, de carácter nacional o internacional, interesados en realizar la evaluación de resultados de los programas gubernamentales sujetos a Reglas de Operación en el ejercicio fiscal 2002”, *Diario Oficial de la Federación*, segunda sección, pp 4-7.
- Skoufias, Emmanuel, “Importancia de la evaluación de los programas sociales: el ejemplo de Oportunidades”, conferencia impartida en el Instituto Tecnológico Autónomo de México, México, 25 de abril de 2003.
- Solarte Pazos, Leonardo (2002), “La evaluación de programas sociales en el Estado neoliberal”, *VII Congreso Internacional del CLAD sobre la Reforma del Estado y la Administración Pública*, Lisboa, 8-11 de octubre.
- Spink, Peter (2001), “Modernización de la administración pública en América Latina: evaluación de los recientes cambios en la gerencia de los ministerios y de las agencias reguladoras”, en CLAD, Serie Estudios de Casos, *Nueva gestión pública y regulación en América latina. Balances y desafíos* [[www.clad.org.ve](http://www.clad.org.ve)], fecha de consulta: 4 de octubre de 2004.
- Spitz, Pierre (2002), “Learning and/or measuring performance? Competing priorities within the UN system”, ponencia presentada en la *V Conferencia Europea de Evaluación*, Sevilla [[www.europeanevaluation.org](http://www.europeanevaluation.org)].
- Stufflebean, Daniel (1974), “Meta-evaluation”, *Ocasional Paper*, núm. 3 [[www.wmich.edu](http://www.wmich.edu)], fecha de consulta: 2 de mayo de 2003.
- Székely Pardo, Miguel (1990), “Identificación de la pobreza global y rural en México: un primer paso para incrementar el grado de acierto de los programas públicos”, tesis para obtener el grado de maestro en políticas públicas, ITAM.
- (diciembre de 1993), “El Programa Nacional de Solidaridad en México”, *Documento de trabajo*, núm. 384, OIT.
- (2003), “Lo que dicen los pobres”, *Cuadernos de Desarrollo Humano*, núm. 13, Sedesol, México.
- Thoenig, Jean-Claude (junio 1996), Conferencia sobre Evaluación, CIDE, México.

## BIBLIOGRAFÍA

- (noviembre 2002), “Upgrading knowledge in public policy and administration”, ponencia presentada en el Seminario Internacional *El estado el arte en administración y políticas públicas*, El Colegio de México, México.
- (2003), presentación titulada “Evaluation of National Public Policies. Learning from experience”, documento en Power Point.
- Toulemonde, Jacques (2000), “Key lessons from the European-wide evaluation practice in the field of social and economics development”, ponencia presentada en la *IV Conferencia Europea de Evaluación*, Lausanne [[www.europeanevaluation.org](http://www.europeanevaluation.org)].
- (2002), “Assuring the quality of evaluations: lessons from European experience”, ponencia presentada en la *V Conferencia Europea de Evaluación*, Sevilla [[www.europeanevaluation.org](http://www.europeanevaluation.org)].
- Uvalle, Ricardo (octubre, 2002), ponencia “Los fundamentos institucionales de la gestión pública”, CD del *VII Congreso internacional del CLAD sobre la reforma del Estado*.
- (octubre, 2003), “La democratización de la administración pública”, CD del *VIII Congreso internacional del CLAD sobre la reforma del Estado y la administración pública*, Panamá.
- Weissbluth, M. (2004), *Seminario de Gerencia Pública*, sede del BID, Estados Unidos.



## Glosario

AECA: Asociación Española de Contabilidad y Administración de Empresas

Aeval: Agencia Gubernamental para la Evaluación de Políticas Públicas y la Calidad de los Servicios

AGEV: Área de Gestión y Evaluación (Uruguay)

AL: América Latina

ASF: Auditoría Superior de la Federación

BID: Banco Interamericano de Desarrollo

BM: Banco Mundial

CAP: Comisión de Administración Pública

C/B: Costo/Beneficio

CCE: Consejo Científico de Evaluación

C/E: Costo/Eficacia

CEPAL: Comisión Económica para América Latina

CERTU: Centre d'Études sur les Réseaux de Transport et l'Urbanisme

CIDE: Centro de Investigación y Docencia Económicas

CIME: Comité Interministerial de Evaluación

CMA: Comisión de Seguimiento y Evaluación (Brasil)

CMH: Contaduría Mayor de Hacienda  
CNDH: Comisión Nacional de Derechos Humanos  
CNE: Consejo Nacional de Evaluación de Políticas Públicas (Francia)  
CNE: Consejo Nacional de Evaluación de la Política de Desarrollo Social (México)  
Colmex: Colegio de México  
Conasupo: Comisión Nacional de Subsistencias Populares  
Comprogrsa: Coordinación Nacional de Progresas  
Coplamar: Coordinación del Plan General de Zonas Deprimidas y Grupos Marginados  
CPyCP: Comisión de Presupuesto y Cuenta Pública  
Diconsa: Distribuidora Conasupo  
EES: European Evaluation Society  
ELECTRA: Élection et choix traduisant la réalité  
ENIGH: Encuesta Nacional de Ingresos y Gastos de los Hogares  
Encel: Encuesta de Evaluación en Localidades  
EUA: Estados Unidos de América  
E&M: Evaluación y Monitoreo  
FMI: Fondo Monetario Internacional  
FNDE: Fondo Nacional para el Desarrollo de la Evaluación  
Fonaes: Fondo Nacional de Empresas en Solidaridad  
GAO: General Accounting Office  
GAPP: Grupo de Análisis de Políticas Públicas  
IAWG: Inter-Agency Working Group on Evaluation  
IFAI: Instituto Federal de Acceso a la Información Pública  
IFPRI: International Food Policy Research Institute  
ILPES: Instituto Latinoamericano de Planificación Económica y Social  
IMSS: Instituto Mexicano del Seguro Social

## GLOSARIO

Indesol: Instituto Nacional de Desarrollo Social

INEGI: Instituto Nacional de Estadística, Geografía e Informática

Infonavit: Instituto Nacional del Fondo para la Vivienda de los Trabajadores

Insol: Instituto Nacional de Solidaridad

INSP: Instituto Nacional de Salud Pública

ISSSTE: Instituto de Seguridad y Servicios Sociales para los Trabajadores del Estado

LGDS: Ley General de Desarrollo Social

Liconsal: Leche Industrializada Conasupo

MCE: Misión de Evaluación y Control del Gasto Público

MCE: Módulo de Control y Evaluación Presupuestaria (Paraguay)

Mideplan: Ministerio de Planificación Nacional y Política Económica (Costa Rica)

Mideplan: Ministerio de Planificación y Cooperación (Chile)

NAO: National Audit Office

NPM: New Public Management

OCDE: Organización para la Cooperación y el Desarrollo Económicos

OEA: Organización de los Estados Americanos

OIT: Organización Internacional del Trabajo

OMB: Office of Management and Budget

OMS: Organización Mundial de la Salud

ONG: Organización No Gubernamental

ONU: Organización de las Naciones Unidas

OPP: Oficina de Planeamiento y Presupuesto (Uruguay)

OSC: Organizaciones de la Sociedad Civil

PAC: Plan de Aceleración del Crecimiento (Brasil)

PEF: Presupuesto de Egresos de la Federación

Pider: Programa de Inversiones Públicas para el Desarrollo Rural

PND: Plan Nacional de Desarrollo

PNDS: Programa Nacional de Desarrollo Social

PMG: Programas de Mejoramiento de Gestión

PNUD: Programa de Naciones Unidas para el Desarrollo

PPA: Plan Plurianual (Brasil)

PPBS: Planning Programming Budgeting System

PRI: Partido Revolucionario Institucional

Progesa: Programa de Educación, Salud y Alimentación

Promap: Programa para la Modernización de la Administración Pública Federal

Pronasol: Programa Nacional de Solidaridad

PRSP: Poverty Reduction Strategy Paper

RCB: Rationalisation des Choix Budgetaires

Sagarpa: Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación

SAM: Sistema Alimentario Mexicano

SCG: Sistema de Control de Gestión (Chile)

SE: Sistema Nacional de Monitoreo y Evaluación (Bolivia)

Secodam: Secretaría de la Contraloría y el Desarrollo Administrativo

Secogef: Secretaría General de la Federación

SED: Sistema de Evaluación del Desempeño (México)

Sedesol: Secretaría de Desarrollo Social

SEGPR: Sistema de Evaluación de la Gestión Pública por Resultados

Semarnat: Secretaría de Medio Ambiente y Recursos Naturales

SEV: Sistema Informático de Evaluación de Gestión Pública (Uruguay)

SFE: Sociedad Francesa de Evaluación

SFP: Secretaría de la Función Pública

SGPR: Sistema de Gerencia Basado en Gestión por Resultados (Honduras)

## GLOSARIO

SHCP: Secretaría de Hacienda y Crédito Público

Siempro: Sistema de Información, Monitoreo y Evaluación de Programas Sociales (Argentina)

Sierp: Sistema de Información de la Estrategia para la Reducción de la Pobreza (Honduras)

SIGFE: Sistema de Información para la Gestión Financiera del Estado

SMMP: Sistema de Seguimiento de Metas Presidenciales (Brasil)

Simep, Sistema de Metas Presidenciales (México)

Sinasid: Sistema Nacional de Seguimiento a Indicadores de Desarrollo (Nicaragua)

Sine: Sistema Nacional de Evaluación (Costa Rica)

Sinergia: Sistema Nacional de Evaluación de Resultados de la Gestión (Colombia)

Sisep: Sistema de Evaluación y Seguimiento de Proyectos (Colombia)

SNPP: Sistema Nacional de Planeación Participativa

SPG: Sistema de Seguimiento de la Programación Gubernamental (Chile)

SPP: Secretaría de Programación y Presupuesto

Ssa: Secretaría de Salud

SSEEP: Sistema de Seguimiento de la Ejecución Presupuestaria (Argentina)

SSEGP: Sistema de Seguimiento y Evaluación del Gasto Público (Perú)

SSEPPS: Sistema de Seguimiento y Evaluación de las Políticas Sociales (Brasil)

TIR: Tasa interna de retorno

NU: Naciones Unidas

UAM-X: Universidad Autónoma Metropolitana, Unidad Xochimilco

UMA: Unidades de Seguimiento y Evaluación (Brasil)

UNAM: Universidad Nacional Autónoma de México

VAN: Valor actual neto



*Evaluación y metaevaluación en las políticas y programas públicos. Estado del arte*, de Myriam Cardozo Brum, número de la Colección Teoría y análisis de la DCSH de la UAM-Xochimilco, terminó de imprimirse el 28 de marzo de 2012, la producción estuvo al cuidado de Váksu editores, Tenorios 222-24-202, colonia Ex Hacienda Coapa, 14300, México, Distrito Federal, 4623 9053 y 5594 9341, vaksu\_entrepalabras@yahoo.com.mx. La edición consta de 1 000 ejemplares más sobrantes para reposición.





## Colección Teoría y Análisis



Casa abierta al tiempo

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades

