

Estado del arte de la evaluación de políticas y programas públicos en México

Myriam Cardozo Brum *

Resumen

México incorpora sistemáticamente los procesos de evaluación al inicio del siglo XXI. En 2006 su organización es asumida por un Consejo Nacional de Evaluación. Dos años después el Gobierno del Distrito Federal (GDF) crea el primer órgano local de evaluación.

La profesionalización de la evaluación es imprescindible para avanzar en este proceso, pero México solo ha contado con algunos cursos y diplomados de formación y la publicación de pocos libros; y no dispone de revistas especializadas en el tema, ni de asociaciones de profesionales que generen espacios de actualización y discusión.

El sistema nacional de evaluación ha realizado avances (e.g. evaluar el diseño de programas), ha mantenido algunos problemas (e.g. el pago de las evaluaciones lo hace el mismo programa evaluado), y también ha incrementado otros (e.g. simplificación de los trabajos). Hoy el número de evaluaciones ha crecido aceleradamente, poniendo en entredicho su integridad, calidad y utilidad, debido a que: a) solo se evalúa parte de la gestión del programa, b) la calidad es afectada, entre otras, por la falta de convocatorias con requisitos específicos para cada caso o de verificación en campo de la información institucional, y c) la utilización real de sus resultados alcanza niveles muy bajos.

* Doctora en Ciencias Políticas y Sociales por la Universidad Nacional Autónoma de México (UNAM) e investigadora de la Universidad Autónoma Metropolitana (UAM), México, DF.

Código de referato: SP.188.XXXIV/15

El sistema del GDF, por el contrario, ha realizado un número reducido de evaluaciones tratando de profundizar en la explicación de los problemas y la adopción de recomendaciones de fondo; sin embargo, su trabajo es incipiente y aún vulnerable si no recibe el apoyo necesario para su consolidación y la efectiva utilización de sus recomendaciones.

Se concluye que crecer aceleradamente puede ser un riesgo para la calidad del sistema, que este debe supervisarse con metaevaluaciones periódicas y que en el corto plazo la utilización de los buenos resultados debe promoverse, entre otros, mediante la introducción de sistemas de incentivos, económicos y simbólicos, individuales y colectivos.

Palabras clave: evaluación – institucionalización – políticas públicas – programas públicos – México

Abstract

Mexico systematically incorporated the processes of evaluation at the beginning of the 21st century. In 2006 their organization was under taken up over by the new National Council of evaluation. Two years later the Government of the Distrito Federal (GDF) creates the first local organization for evaluation.

The professionalization of evaluation is essential, but Mexico only had a few courses and diplomas and the publication of a few books. There are no journals or professional associations that generate new information and debate.

The national system of evaluation has made some progress (e.g. assessing program design), has kept old problems (e.g. payment of assessments are still made by the evaluated program), and has magnified other existing problems (e.g. oversimplification of the work). Today the number of evaluations has grown dramatically calling into question its integrity, quality and usefulness, since: a) only part of the management of the program is evaluated, b) the quality is affected, among others, by the lack of summons with specific requirements to each case, or by the lack of verification of the institutional information in the field, c) There are very low levels of utilization of their results.

On the other hand, the system of the GDF has made a small number of evaluations trying to improve problem definition and the follow up and adoption of recommendations. However, their work is incipient and is still vulnerable if it does not receive support for its consolidation and effective use of its recommendations.

It is concluded that growing rapidly may be a risk for the quality of the system, which should be monitored with periodic meta-evaluations. Also, that in the short-term, the utilization of the good results must be promoted, among others, by introducing systems of incentives, both economic and symbolic, individual and collectives.

Keywords: evaluation – institutionalization – public policies – public programs – México

Introducción

MÉXICO se incorpora de manera sistemática a los procesos de evaluación de políticas y programas públicos al inicio del siglo XXI. El número de sus trabajos ha crecido aceleradamente y los tipos de evaluación se han diversificado desde la creación, a nivel federal, del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) en 2006, y, en menor grado y con carácter local, del Consejo de Evaluación del Desarrollo Social del Distrito Federal (Evalúa DF). Ambos Consejos han organizado en forma diferente sus sistemas de evaluación e impulsado enfoques teóricos alternativos.

A continuación se presentan los principales antecedentes de este proceso para luego analizar el estado actual del arte, incluyendo su institucionalización y profesionalización, su normatividad, organización, planeación y ejecución, así como los mecanismos enfocados a mejorar la calidad de los trabajos y a incentivar su utilización. La información analizada en el artículo se ha ido generando dentro de un proyecto de investigación desarrollado a nivel universitario ¹ desde el año 2003.

1. Breves antecedentes e institucionalización formal

En 1976 se creó la Secretaría de Programación y Presupuesto (SPP), entonces 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, que debía diseñar e implantar un sistema nacional de evaluación; sin embargo, acabó realizando estudios parciales y de poco impacto en el proceso.

En 1982 inició funciones la entonces Secretaría de la Contraloría General de la Federación (Secogef), encargada de la fiscalización, el control interno y la evaluación de la administración pública. Al siguiente año la SPP le transfirió completamente las funciones de evaluación pero la Secogef no pudo cumplirlas debido a la crisis económica que se atravesaba.

Unos pocos años después (aproximadamente por 1989) comenzaron a aplicarse algunas de las ideas del *New Public Management* como consolidar una

¹ Universidad Autónoma Metropolitana.

cultura administrativa que propiciara la eficiencia y la productividad en la gestión de las instituciones gubernamentales, promoviera la evaluación de la calidad y la transparencia en la prestación de los servicios públicos y previniera y combatiera la corrupción.

Dicho enfoque se reforzó en 1995 con un programa de modernización que identificaba entre los principales problemas a resolver: las deficiencias en los mecanismos de medición y evaluación del desempeño de la gestión pública, para los que se proponía el diseño de un sistema basado en objetivos claros y metas cuantificables, indicadores de desempeño, revisión y adaptación del Sistema Integral de Información y fortalecimiento y mayor independencia de los órganos internos de control preventivo y correctivo. En los años siguientes la Secretaría de Hacienda y Crédito Público (SHCP) fue encargada de estudiar la posibilidad de reformar el sistema presupuestal, vinculándolo a la evaluación de resultados de la gestión pública, pero esto solo se implementó parcialmente.

Si bien desde 1986 hubo intentos de diseñar metodologías y de evaluar programas, fue en la segunda mitad de los noventa que se realizaron los primeros trabajos, iniciando en 1997 el proyecto de evaluación externa más ambicioso solicitado por el gobierno mexicano, el del Programa de Educación, Salud y Alimentación.

El proceso de institucionalización formal de la evaluación inicia en el nuevo siglo con la decisión de la Cámara de Diputados de exigir la evaluación externa anual de los programas financiados con recursos fiscales. Destaca el trabajo realizado por la Secretaría de Desarrollo Social (Sedesol) que creó una Subsecretaría encargada de la evaluación de los 20-30 programas que se encontraban bajo su responsabilidad. También la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Sagarpa) que realizó evaluaciones mediante convenio con la *Food and Agriculture Organization of the United Nations* (FAO). En 2002, se estableció el Sistema de Seguimiento de Metas Presidenciales, bajo la responsabilidad de la entonces Secretaría de la Contraloría y el Desarrollo Administrativo, para uso interno del Gobierno Federal y, al año siguiente, se aprobó la Ley General de Desarrollo Social (LGDS) que previó la creación de un Consejo Nacional de Evaluación, que recién sería creado en 2006 como Coneval y se complementaría en 2007 con el Sistema de Evaluación del Desempeño de la SHCP, basado en la Ley de Presupuesto y Responsabilidad Hacendaria aprobada un año antes. Estas decisiones consolidan la institucionalización formal de los procesos de evaluación a nivel federal.

A nivel local, ante la casi inexistencia de un proceso de evaluación en la administración del Gobierno del Distrito Federal (GDF) que contribuyera a mejorar la gestión de sus políticas y programas sociales, en 2007, su Jefe de Gobierno creó Evalúa DF, el que fue instalado al siguiente año. Además, en el marco del Modelo de Gestión para Resultados, la Coordinación General de Modernización Administrativa del GDF desarrolló el Programa de Monitoreo y Evaluación del Desempeño Gubernamental 2010-2012, destinado a evaluar sus instituciones a través de sistemas de indicadores, estadísticas y encuestas, y emitir recomendaciones basadas en información objetiva y confiable.

2. Profesionalización de la evaluación

En materia de formación, México no ha contado hasta ahora con programas académicos especializados en el tema, solo algunos pocos cursos en posgrados más amplios.² Si bien fueron las universidades y otras instituciones educativas y de investigación las que iniciaron la realización de evaluaciones independientes y han estado desarrollando trabajos por consultoría, a la fecha no han existido programas especialmente enfocados al desarrollo de esta incipiente profesión y puede constatarse el reducido número de tesis vinculadas al tema, que en los últimos años se han venido incrementado en forma acelerada. Además, existe una fuerte 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 los Estados Unidos de América (EUA) y, en mucho menor medida, de Europa o Canadá; respondiendo a contextos muy diferentes de los nuestros.

Los esfuerzos realizados en este campo incluyen, entre otros: un primer diplomado en evaluación organizado por el Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria de la Cámara de Diputados y otro promovido por la Universidad Autónoma Metropolitana, 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 Distrito Federal. Por otra parte, Coneval junto con la Comisión Económica para América

² Maestría en Políticas Públicas de la Universidad Autónoma Metropolitana, maestrías en Evaluación Educativa de la Universidad Nacional Autónoma de México y en Evaluación y Administración de Proyectos de la Universidad Panamericana.

Latina (CEPAL) y, por otro lado, la SHCP, han capacitado intensamente en torno al diseño del marco lógico. La SHCP (2012) informó que de marzo 2009 a diciembre 2011 instrumentó un curso de educación a distancia sobre la MIR (Matriz de Indicadores para Resultados) que atendió a 2 459 servidores públicos.

Recientemente se ha instalado también una sede del *Center for Learning on Evaluation and Result* (CLEAR) en el Centro de Investigación y Docencia Económicas AC (CIDE), encargado de proporcionar conocimiento institucional y técnico de vanguardia a sus clientes de América Latina para

“fortalecer las capacidades de evaluación, monitoreo y gestión por resultados; promover la utilización de la evidencia empírica como la herramienta principal para la evaluación y para la rendición de cuentas; y fortalecer el liderazgo de América Latina en estas áreas mediante una estrecha vinculación con los otros Centros regionales y a través de la generación de espacios para el aprendizaje entre regiones” (CLEAR, s/f).

Hace poco más de un año inició la impartición de su primer Diplomado de Políticas Públicas y Evaluación y actualmente está programando el cuarto para marzo de 2015, con un costo de USD 2 200, bastante elevado para el contexto latinoamericano.

Con apoyo de CLEAR y la planta de investigadores del Centro de Investigación en Evaluación y Encuestas del Instituto Nacional de Salud Pública (INSP), también se aprobó en 2013 un programa virtual de Especialidad de un año de duración en Evaluación Integral de Programas y Políticas de Desarrollo Social. Busca formar profesionales capaces de abordar la evaluación de un programa o una política de desarrollo social desde una perspectiva integral. Está dirigido a tomadores de decisión, operadores de programas, académicos, consultores o profesionales interesados en formarse en el área de evaluación y/o que actualmente formen parte activa de programas sociales. Trata de capacitar en la realización de evaluaciones para las distintas etapas del proceso de políticas y programas, tanto con metodología cualitativa como cuantitativa. Tiene un costo total de USD 3 150.

Coneval se ha mostrado más preocupado por la capacitación continua de quienes están realizando las evaluaciones que por la formación en general, por lo que acaba de realizar (octubre de 2014) el Segundo Seminario dirigido a las Entidades Federativas del país en materia de “Implementación de un sistema de monitoreo, evaluación y uso de la medición de la pobre-

za para la toma de decisiones”. Además de los temas del título se incluyen otros vinculados a la planeación como el uso de información para la creación de programas sociales.

A nivel municipal, las instituciones convocantes de la Conferencia Anual del Municipio organizaron el pasado mes de noviembre un evento de medio día sobre Medición y Evaluación del Desempeño en la Administración Pública, en el marco del Foro Sistema de Evaluación del Desempeño (en gobiernos locales) para Presupuesto basado en Resultados. La intención es que los municipios vayan preparando su propio sistema, que deberá estar listo a fines de 2015.

Por otra parte, México no ha contado con asociaciones de profesionales independientes, salvo la Asociación Mexicana de Profesionales de la Evaluación A.C. que reúne principalmente a especialistas del desarrollo rural, por lo que no representa a todos los sectores ni tiene cobertura nacional. A la fecha ha organizado dos congresos, el tercero que tenía programado en 2013 debió cancelarlo. Recientemente, se identificó también una red denominada Academia Nacional de Evaluadores de México. Por otro lado, muchos evaluadores participan en los congresos y publicaciones virtuales de la Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLac) y de la Red Latinoamericana y del Caribe de Monitoreo y Evaluación (Redlacme), impulsada por el Banco Mundial (BM), el Banco Interamericano de Desarrollo (BID) y el Coneval, que cuenta con un capítulo mexicano (Red Mexicana de Monitoreo y Evaluación, Redmexme).

Finalmente, México tampoco dispone de revistas especializadas; los artículos en el tema son poco numerosos y han sido publicados en revistas como *Andamios*, *Alegatos*, *Política y Cultura*, *Gestión y Política Pública*, *Investigación Ambiental*, *Convergencia*, y *Estudios Sociales*. En cuanto a los libros se dispone de alrededor de 15 publicados a lo largo de 10 años.

En síntesis, se han desarrollado esfuerzos importantes aunque aún insuficientes para formar evaluadores, que no se han acompañado de espacios para la discusión y publicación.

3. El sistema federal de evaluación

Ante la falta de experiencia previa y procesos de capacitación *ad hoc*, en años inmediatos al 2001, las evaluaciones externas presentaban desafíos importantes como mejorar la claridad de las normas que guiaban su con-

ducción, contar con un fondo independiente de recursos para la contratación de evaluadores al margen del programa evaluado, seleccionar siempre al equipo evaluador mediante concurso, avanzar en la realización de evaluaciones participativas centradas en el aprendizaje de funcionarios y ciudadanos, profesionalizar el campo, y realizar un mayor esfuerzo por institucionalizar la cultura de la evaluación (Cardozo, 2012a).

A partir de 2003, la LGDS prevé la realización de evaluaciones con la finalidad de revisar periódicamente el cumplimiento del objetivo social de los programas, metas y acciones de la política de desarrollo social, para corregirlos, modificarlos, adicionarlos, reorientarlos o suspenderlos total o parcialmente (art. 72). Señala que para efectuar su evaluación, los programas sociales deberán incluir los indicadores de resultados, gestión y servicios para medir su cobertura, calidad e impacto. Los primeros miden el cumplimiento de los objetivos sociales y los otros dos, los procedimientos y calidad de los servicios (art. 74-76). La ley enfatiza la importancia de la participación social y ciudadana mediante el mecanismo de contralorías sociales (art. 61-71) y determina que la evaluación será anual, aunque también podrá ser multianual en algunos casos (art.78).

El Reglamento de la ley anterior establece en su art. 76 que el Consejo Nacional de Evaluación definirá los criterios para la elaboración de los indicadores antes referidos y en los trabajos de evaluación se deberán observar los lineamientos que este establezca (art. 77). Por su parte, el art. 79 aclara que las evaluaciones de impacto de los programas de desarrollo social podrán realizarse a través de procesos anuales o multianuales, de conformidad con la información y naturaleza del programa, lo que constituye un avance frente a la normatividad previa que prescribía su realización anual.

Para efectuar una evaluación integral se tendrían que cubrir también los antecedentes del problema y política o programa evaluado, el enfoque teórico que lo avala dentro del conjunto de los posibles y la situación económica, social y política en el marco de la cual se evalúa. Sin embargo, la normatividad vigente no los solicita. La LGDS y su reglamento, no obstante, exigen evaluaciones anuales de: 1) el logro de objetivos y metas (eficacia), 2) la gestión realizada (procesos) y 3) la cobertura y calidad de los servicios (resultados). Esto implica que: 1) el proceso de formulación y diseño del programa puede no ser revisado, como estuvo sucediendo en el período 2001-2006; 2) tampoco se menciona el análisis de los recursos (aunque puede ser incluido en la investigación sobre la gestión) cuya oportunidad, calidad y buena utilización permitirían, conjuntamente con

los resultados, evaluar la eficiencia; 3) no incluye un estudio del contexto económico y político en el marco del cual tiene sentido valorar el programa y los resultados alcanzados en el año y 4) no solicita la explicitación del enfoque teórico en que se sustentará la evaluación.

En 2007, Coneval, conjuntamente con dos secretarías vinculadas al control de la administración pública (SHCP, y Secretaría de la Función Pública, SFP), comenzó a realizar un Programa Anual de Evaluaciones (PAE) que incluyó todos los programas con contenido social a cargo de la Sedesol pero también de otras secretarías, con lo cual su número aumentó en forma muy considerable, llegando casi a 200 en 2009 frente a las 20-30 del período previo (ver cuadro N° 2). También fueron publicados los nuevos “Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal” (en adelante Lineamientos) para guiar la realización de evaluaciones externas que sustituirían al Acuerdo que había estado vigente desde 2002 para normar esta actividad.

Los tipos de evaluaciones definidos por los Lineamientos fueron: estratégicas y de programas federales (Cuadro N° 1). Las primeras, como su nombre lo indica, se aplican a las estrategias, políticas e instituciones relacionadas con un programa o conjunto de programas, mientras que las demás se dirigen a los programas concretos y, a su vez, se componen de:

a) Evaluaciones de consistencia y resultados (ECR) referidas al diseño, desempeño global y resultados basados en el seguimiento de indicadores, mediante trabajo de gabinete. Cuentan con TR genéricos, originalmente de 100 preguntas, que luego fueron reducidas a 51. Dentro de estas se desagregó otro tipo de evaluaciones referidas exclusivamente a diseño. También disponen de TR genéricos, simplificados de 34 a 21 preguntas, que acaban de aumentarse a 30. Retoman un instrumento de planeación del marco lógico utilizado por varios organismos internacionales, al que se denomina Matriz de Indicadores para Resultados (MIR).

b) Evaluaciones de indicadores, que analizan su pertinencia y alcance para la medición de resultados, incluyendo trabajo de campo. No se ha realizado ninguna evaluación de este tipo ni se ha diseñado una metodología específica pero se puede considerar que, en parte, se retoman en las evaluaciones de diseño.

c) Evaluaciones de procesos, que revisan su eficacia y eficiencia operativas y su contribución al mejoramiento de la gestión. Cuentan con TR genéricos desde 2009 en que empezaron a aplicarse, los que fueron modificados en 2013. A la fecha solo se han realizado 27 evaluaciones de este

tipo; sin embargo, también pueden quedar incluidas cuando se analiza la operación en las ECR.

d) Evaluaciones de impacto, que miden cambios en los indicadores de resultados atribuibles a la gestión del programa, con base en metodologías revisadas por la Sedesol, la SFP y el Coneval.

e) Evaluaciones específicas, que incluyen todas las no comprendidas en los tipos previos, por lo que pueden implicar tanto trabajo de gabinete como de campo.

Este es el conjunto previsto en los lineamientos que se estuvo aplicando regularmente desde 2007. Posteriormente se fueron desagregando otros tipos (ver Cuadro N° 1):

f) Las evaluaciones específicas de desempeño (EED), que Coneval comenzó a contratar en 2008 como parte de las evaluaciones específicas y para las que recién elaboró un modelo de TR en la generación 2012-2013, que solicita: datos generales, resultados/productos, evolución de la cobertura, seguimiento a recomendaciones y datos de la instancia evaluadora. Dicho modelo expresa que “se debe realizar mediante trabajo de gabinete y únicamente con base en la información proporcionada por Coneval” (Coneval, 2013: 7), con lo que prácticamente se prohíbe usar información de fuentes alternativas, lo que constituye una restricción grave a la libertad del evaluador.

f) Las evaluaciones complementarias, también iniciadas en 2008, son consideradas de aplicación opcional, de acuerdo con las necesidades e intereses de las dependencias y entidades, con el fin de mejorar su gestión y obtener evidencia adicional sobre su desempeño; sin embargo, requieren la opinión del Coneval sobre sus TR.

h) Las integrales de desempeño, que Coneval agregó en 2009, tampoco cuentan con una metodología explícita, tienen el propósito de contribuir a mejorar el diseño de políticas públicas y de facilitar una visión de la interacción que existe entre distintos programas sociales federales que aportan al mismo tema: salud, asistencia social, educación, empleo, desarrollo forestal, población indígena y vivienda.

i) Las de diagnóstico, cuyos elementos mínimos fueron establecidos hace ya varios años, pero recién en 2014 se inició su aplicación con el objetivo de analizar un problema concreto, sus causas y consecuencias, el presupuesto requerido, así como las poblaciones involucradas y realizar una propuesta para atenderlo, con su correspondiente diseño, a la luz de experiencias internacionales exitosas.

Cuadro N° 1. Características por tipo de evaluación

Tipos	Incluidas en Lineamientos	Con metodología expresa	Con TR	Trabajo de campo explícitamente incluido
Estratégicas	X	X		
ECR	X	X	X	
Indicadores	X			X
Procesos	X	X	X	X
Impacto	X	X		Posible
Específica	X	X		Posible
Diseño	X	X	X	
Complementarias	X	X		
Integrales de Desempeño				
Específicas de Desempeño		X	X	
Diagnóstico	X	X		

Fuente: elaboración propia con base en Coneval-SHCP-SFP (2007) y Coneval (s/f)

Como puede verse, se trata de un conjunto de tipos no siempre excluyentes ni claramente diferenciados, algunos con TR estandarizados, la mayoría son trabajos exclusivamente de gabinete y los que incluyen trabajos de campo se han aplicado muy poco. Si esta situación se contrasta con los contenidos exigidos por la LGDS y su reglamento, el tipo más completo corresponde a la ECR, que incluso agrega la necesaria evaluación de diseño y de la planeación estratégica, por lo que debería constituir la evaluación anual fundamental; sin embargo, solo se ha aplicado en 2007 y 2011. La realización de evaluaciones complementarias, estratégicas, específicas, de desempeño o de diseño puede ser muy útil pero no cubre todos los aspectos exigidos anualmente por la LGDS.

Si bien el BM (s/f) elaboró un documento de análisis de los Lineamientos, en realidad no analiza los tipos de evaluaciones a aplicar sino que informa sobre los avances emprendidos por el Consejo en materia de difusión de resultados, impulso a la capacitación de evaluadores, mejora de sistemas de información, revisión de las leyes que regulan los contratos, etc.

Cuadro N° 2. **Evaluaciones federales realizadas en 2007-2011**

Tipo de Evaluación	2007	2008	2009	2010	2011	Total
Diseño	14	19	33	17	25	108
ECR	111	0	0	0	0**	111
Procesos	0	0	5	8	1	14
Específicas	6	18	22	1	0	47
EED	0	131	132	133	0	396
Impacto	3	8	4	1	0	16
Estratégicas	2	1	1	1	0	5
TOTAL	136	177	197	161	26	697
Complementarias*	0	5	5	4	4	18

Notas: * Las evaluaciones complementarias no se incluyen en el PAE.

** A la fecha en que se elaboró el documento usado como fuente aún no se acababan las ECR realizadas en 2011.

Fuente: adaptado y corregido de SHCP (2012).

En la práctica, en 2007-2011 se realizaron 697 evaluaciones incluidas en el PAE (81 % de las planeadas) y 18 complementarias. Las evaluaciones de diseño de nuevos programas se han aplicado con regularidad. El resto de ellas se ha concentrado en las EED, que son las más simples, mientras que las de impacto, las más complejas, alcanzaban apenas a 16 (Cuadro N° 2). Recientemente, en 2014, se han publicado como evaluaciones específicas las Fichas de monitoreo y evaluación, una especie de informe muy limitado (formato de dos páginas), realizado por el propio órgano operador. Si esta situación se compara con la que estuvo vigente en 2001-2006, se encuentra que en aquel período las evaluaciones trataban de ser integrales (si bien no incluían el diseño) y sus TR, aunque muy similares, recogían algunas especificidades de cada caso, por lo que a partir de 2007 se identifica una clara tendencia hacia la estandarización y la simplificación. Los cambios introducidos a los TR de las ECR y de diseño refuerzan esta afirmación.

En algunos casos se hacen estudios de mercado, previos a la contratación, solicitando cotizaciones para definir el precio a pagar por parte de los propios programas que serán evaluados (salvo las contratadas directamente por Coneval). La información sobre los costos de los trabajos ha sido siempre difícil de obtener, pero en el período 2007-2014 se han encontrado casos que varían entre USD 3 995 para una evaluación específica en 2009-2010 hasta USD 2'578.750 para una de impacto (de Oportunidades) en 2008-2009, con variaciones algunas veces difíciles de justificar. Si se

considera la evolución de costos mostrada en el cuadro N° 3 en casos seleccionados se concluye que en la actualidad la mayoría de las realizadas están muy por debajo del precio pagado en el período 2001-2006, fundamentalmente debido a la desaparición del trabajo de campo, lo que marca una tendencia hacia el abaratamiento.

Cuadro N° 3: **Costos de las evaluaciones**

Año	Evaluaciones consideradas	Monto promedio (en USD)
2002	31 evaluaciones realizadas por Sedesol	175.465
2005	4 evaluaciones contratadas por Sedesol	36.718
2007	5 evaluaciones de diseño solicitadas por Coneval	23.975
2008	16 ECR	49.780

Fuente: Cardozo (2009).

Un tema muy cercano es el de los mecanismos de contratación. En 2001-2006 se utilizaba principalmente la invitación a concursar a por lo menos tres instituciones, actualmente pocos son los casos en que esto se realiza, sustituyéndolo por adjudicaciones directas, no solamente a instituciones de educación superior, de investigación científica u organizaciones no lucrativas como marca la LGDS (art. 73) sino también a empresas consultoras.

En síntesis, la constitución del Coneval en 2006 significó un gran paso adelante en la mejor organización del sistema de evaluaciones externas (un programa anual de evaluaciones, lineamientos más precisos, inclusión de la evaluación de diseño, etc.); sin embargo, se han mantenido muchas de las debilidades y desafíos identificados en las metaevaluaciones previas realizadas (Cardozo, 2012a), como por ejemplo, la disponibilidad de un fondo independiente para contrataciones o la realización de evaluaciones participativas.

En 2010, el BID publicó su Índice de Gestión para Resultados (García y García, 2010) incluyendo a 25 países de América Latina y el Caribe, uno de cuyos cinco pilares está constituido por el sistema de monitoreo y evaluación: México ocupó el cuarto lugar, después de Chile, Brasil y Colombia, todos clasificados en el nivel alto de desarrollo del índice. Se trata de un buen lugar (calificación de 3.2 en la escala de 0 a 5, con una diferencia aproximada de 1.3 respecto de Chile), pero también indica que aún hay retos importantes que superar en materia de desempeño gubernamental.

El Coneval también ha sido objeto de evaluaciones y auditorías, como la del BM en 2008 y la de la Auditoría Superior de la Federación (ASF) de 2010. Más recientemente Szekely (2013) aplicó 24 entrevistas y encuestas a diversos actores relevantes en el sector de la evaluación y la medición de la pobreza para conocer su percepción sobre Coneval. El 48 % de los funcionarios encuestados manifestó estar muy satisfecho y el 39 % algo satisfecho; mientras que en cuanto a su información y análisis, el 56 % y el 31 % los consideró respectivamente algo o muy requeridos para los elaboradores de políticas públicas y los tomadores de decisión. Adicionalmente, los entrevistados identificaron tres retos para dicho Consejo: mantener su independencia del gobierno y la calidad de sus productos, así como lograr que las evaluaciones tengan un mayor impacto en mejorar los programas sociales (asunto ya señalado en la evaluación externa realizada por el BM en 2008).

Otra evaluación (Rojas, 2013) se concentró en el desempeño organizacional de Coneval. Para ello aplicó 26 entrevistas a personas vinculadas con los procesos de evaluación. Encontró que han cambiado las circunstancias externas en relación con el estilo político del gobierno federal (más propenso a los pactos y alianzas, a otorgar plena autonomía al Coneval, a modificar la forma de elegir a sus Consejeros, etc.), la tendencia a poner más atención a la fase de implementación de políticas y programas (más que a su diseño o evaluación *ex post*) y la alineación de las acciones en torno a políticas prioritarias de naturaleza multisectorial (no meramente sectoriales). Estas nuevas expectativas y demandas indujeron al evaluador a proponer que Coneval asuma nuevos objetivos, adicionales a los previstos en la LGDS, como el apoyo a los gobiernos subnacionales en materia de evaluación.

En algunas de sus estrategias Coneval comparte responsabilidades con otras instancias y, según mencionaba Rojas (2013), el gobierno federal estaba en disposición de otorgarle plena autonomía, mientras algunos funcionarios entrevistados por Székely (2013) habían señalado que quizá resultara recomendable otorgarle autonomía operativa, presupuestaria y de decisión dentro de la Administración Pública Federal, de manera tal que no se encontrara sectorizado a la Sedesol, como hasta ese momento. Esto se consumó con la reforma política-electoral, aprobada por el Congreso y publicada en febrero de 2014, que le otorgó el carácter de órgano autónomo con personalidad jurídica y patrimonio propios. Además se previó que esta decisión sería respaldada por la Constitución mediante un nuevo apartado C en su artículo 26. Sin embargo, la conveniencia de esta decisión ha sido discutida tanto en algunos medios académicos como por el propio Consejo (Cejudo, 2014; Coneval, s/f).

Al margen del Coneval, también es necesario mencionar que recientemente han sido creados nuevos órganos federales encargados de la evaluación en temas educativos y ambientales. Además, en paralelo a las actividades de evaluación del Poder Ejecutivo federal, la ASF de la Cámara de Diputados ha ido realizando múltiples trabajos dirigidos a auditar el desempeño de los programas públicos. A nivel de las entidades federativas del país comienzan a crearse algunos órganos locales de evaluación, siendo el más antiguo el del GDF, conformado en 2008, que se analiza en el siguiente punto. Y finalmente, se espera que los propios municipios inicien la instalación de sistemas durante el presente año.

4. El sistema de evaluación del GDF

El Artículo 42 de la Ley de Desarrollo Social del Distrito Federal (LDSDF) y su Reglamento (RLDSDF) establecen la necesidad de mejorar la gestión de sus políticas y programas, para lo cual se deben detectar sus aciertos y fortalezas, identificar sus problemas, recoger ampliamente la opinión de los participantes o beneficiarios, y en su caso, formular las observaciones y recomendaciones para su reorientación y fortalecimiento. Para alcanzar este objetivo, Evalúa DF ha organizado un sistema de evaluación sustantivamente diferente al federal (Cardozo, 2011 y 2012b). A título de ejemplo, su normatividad no lo obliga a evaluar sus programas anualmente y establece que sus recomendaciones tienen carácter vinculatorio.

Ante la ausencia de trabajos evaluativos previos, desde el inicio de su trabajo Evalúa DF impartió diplomados y talleres de formación, publicó una serie de lineamientos y una pequeña guía de orientación que define los tipos de evaluación encarados por este Consejo, tanto de programas como de políticas y de servicios. Cada uno de estos se definen a su vez en función de las etapas del proceso: de formulación o diseño, de implementación u operación, de resultados y de impacto (Cardozo y Mundo, 2012). De esta manera se espera que el evaluador recoja y juzgue evidencias pero también que actúe como facilitador de un proceso de aprendizaje.

Las evaluaciones efectuadas en 2008 contaron con TR genéricos con algunas preguntas específicas por programa, pero desde 2009 estos se diseñan especialmente en función de la política o el servicio a evaluarse. Dichos TR solicitan propuestas que, además de diseñar su metodología y planear sus actividades, incluyan los antecedentes del problema y política o programa en cuestión, el contexto social, político y económico en que se ha desarrollado y un marco teórico para su estudio.

En la práctica, se han realizado 26 trabajos de evaluación de políticas, programas y servicios públicos y dos metaevaluaciones, seleccionados con base en criterios explícitos y financiados con un fondo específico administrado por Evalúa DF para garantizar la independencia del evaluador. A partir de 2009 se pasa de la evaluación de programas a la de políticas, disminuyendo su número en función de su complejidad. La mayoría gira en torno a las primeras etapas del proceso (diseño/formulación y operación/implementación); solo uno de los trabajos ha evaluado el impacto y realizado trabajo de campo (Cuadro N° 4).

Cuadro N° 4. **Evaluaciones realizadas por Evalúa DF 2008-2013**

Tipo de evaluación	2008	2009	2010	2011	2012	2013	Total
Diseño y operación de programas	13						13
Formulación e implementación de políticas		4	2		2	1	9
Diseño, operación y resultados de políticas			1				1
Operación, resultados e impacto de servicios				1			1
De presupuesto enfocado a temas transversales						2	2
Metaevaluaciones		1	1				2
Total	13	5	4	1	2	3	28

Fuente: elaboración propia con base en información de Evalúa DF (s/f).

Las convocatorias de Evalúa DF publican los montos disponibles para la realización de cada trabajo y los equipos se seleccionan mediante concurso dentro de una convocatoria abierta entre quienes cumplieron los requisitos y presentaron la mejor propuesta técnica, con base en criterios preestablecidos.

Aunque evidentemente existen aspectos mejorables en la organización de la evaluación por este Consejo, el trabajo realizado ha constituido un avance relevante, sobre todo si se tiene en cuenta que no había antecedentes en el GDF y que se optó por una organización y un enfoque teórico distinto (cercano al constructivismo) al desarrollado por Coneval, por lo que se tuvo que ir construyendo paso a paso su modo de encarar los trabajos.

5. Calidad de las evaluaciones

La calidad de las evaluaciones depende de la normatividad, la información disponible o la metodología aplicada, pero también de la organización del trabajo (Cardozo, 2006). A nivel federal, en 2001-2006 se habían detectado debilidades metodológicas en los esquemas de muestreo, la información recogida y en la emisión de valoraciones fundamentadas sobre la gestión del programa, pero también problemas de organización y planeación como la falta de evaluaciones de diseño, de impacto plurianual, de conjuntos de programas interrelacionados y la aplicación de metaevaluaciones, entre otros. A partir de 2006 hubo avances, estancamientos y algunos retrocesos que se refieren a continuación, con énfasis en los de tipo organizativo y de planeación, que también afectan la calidad de los informes finales logrados.

Un documento básico para contribuir a la calidad de la evaluación lo constituyen los TR en que se especifica la solicitud. Sin embargo, en materia de desarrollo social federal, un alto porcentaje de los revisados en una muestra no representativa de 16 casos de 2008-2013 (56 %) careció de ellos y el resto fueron genéricos. La ausencia de TR puede constituir parte de la explicación de la limitada contribución de los informes al mejoramiento de los programas.

La convocatoria por invitación a cuando menos tres instituciones se ha reducido notoriamente, sustituyéndose por asignaciones directas o contrataciones por convenio. A pesar de esto, en general, los equipos evaluadores fueron de alto nivel; en cambio, se detectaron problemas en cuanto a su composición disciplinaria (énfasis en economía en detrimento de otras disciplinas, también necesarias para la comprensión del programa). Los procesos metodológicos se centraron en la recolección de información de fuentes institucionales y el análisis documental, como establecen los Lineamientos, pero en los casos de evaluación de procesos y de impacto también se aplicaron encuestas en campo. En el caso de estos últimos, los modelos aplicados fueron de carácter experimental, con el objetivo de mostrar una atribución del tipo “causa-efecto”, los que han sido ampliamente discutidos en la literatura y, en la práctica, solo han logrado parcialmente su finalidad.

Los informes finales están adecuadamente estructurados, su redacción es clara y su tamaño muy variable: desde dos páginas en el caso de una ficha de monitoreo a 315 en una evaluación de impacto. Todos se encuentran disponibles en la página web del Coneval pero, excepto en las evaluaciones específicas recientes, estos no se acompañan de un documento de reacción institucional ante el mismo.

Salvo algunas excepciones, se trata de trabajos de calidad aceptable, con algunos aspectos poco cubiertos y sin una clara correlación con sus costos (cuadro N° 5). Sin embargo, sus solicitudes (TR) se han simplificado tanto que la información agregada es muy limitada en la mayoría de los casos. En el caso de las evaluaciones de impacto, el problema gira en torno al enfoque exigido en los Lineamientos: modelos experimentales, costosos y poco efectivos.

Cuadro N° 5. **Costos disponibles y calidad de las evaluaciones correspondientes**

Costos (en USD)	Calidad no aceptable	Calidad aceptable	Calidad alta	Calidad muy alta
3.995	X			
6.986				X
7.751	X			
26.403			X	
51.207				X
55.429			X	
96.225			X	
278.190		X		
805.153			X	
2'578.750			X	

Fuente: elaboración propia con datos de Cardozo (mimeo).

Si comparamos esta situación con la vivida en 2001-2006 durante la contratación de evaluaciones por la Sedesol (Cardozo 2006 y 2012a), no hay dudas de que el sistema se ha expandido ampliamente, que ha incorporado la evaluación del diseño, y que está mejor regulado por los Lineamientos, a pesar de que los tipos de evaluaciones tomen nombres poco sugerentes de su contenido y algunos ni siquiera se definan. Sin embargo, la riqueza de contenido de las primeras evaluaciones era mucho mayor porque normalmente implicaban un trabajo de campo y los estudios aspiraban a ser integrales, intentando al menos cubrir la ejecución, los resultados y el impacto. Actualmente, muchas evaluaciones (EED) no son más que una especie de informe elaborado por un agente externo pero solamente con información institucional, que no puede verificar. Otras son documentos preparados a partir de evaluaciones previas (e.g. fichas de monitoreo, “metaevaluaciones”). Incluso la extensión del informe era notablemente mayor en el período inicial. El actual aumento del número de evaluaciones se ha logrado a costa de la extensión y la calidad del trabajo, con una notable reducción del costo por evaluación (salvo excepciones, como la de Oportunidades).

La práctica federal actual cumple generalizadamente con el requisito formal de realizar una evaluación anual a través de EED, de diseño y recientemente de diagnóstico, y cada tres o cuatro años realiza una ECR, bastante completas salvo por la falta de medición del impacto. Los demás tipos (de procesos, integrales de desempeño, complementarias, estratégicas, de impacto), que podrían ser más efectivos, se han vuelto excepcionales; especialmente los de impacto.

De las evaluaciones revisadas, pocas son las que contienen recomendaciones novedosas o al menos refuerzan acciones útiles (evaluación de procesos, ECR, estratégicas o de impacto); la mayoría no precisa la manera de llevarlas a cabo y la instancia responsable, sus costos y tiempos, lo que no permite garantizar su viabilidad y probabilidad de ser aceptada por los responsables de los programas correspondientes. También se percibe la recurrencia a recomendar nuevas evaluaciones, especialmente de impacto debido a su notable escasez.

En consecuencia, con el objetivo de mejorar la calidad a nivel federal mediante una mejor organización, hemos propuesto: a) cumplir con la obligación legal de realizar siempre evaluaciones externas; b) emitir siempre TR para cada tipo de evaluación y adecuarlos a la problemática concreta del programa a evaluar; c) conformar equipos interdisciplinarios de evaluación, experimentados y *ad hoc* a cada programa; d) emitir siempre convocatorias, dirigidas al menos a tres instituciones; e) explicitar los criterios de selección de propuestas; f) transparentar los montos contratados; g) dar más importancia a la realización de trabajos de campo que verifiquen y complementen la información institucional; g) intentar evaluaciones de impacto de tipo contributivo; h) exigir un mínimo de extensión, calidad y aportaciones a los informes de evaluación, explicitando y fundamentando mejor las conclusiones, en congruencia con su tipo; e i) informar siempre sobre la respuesta institucional a la evaluación, las recomendaciones y su cumplimiento, desglosadas para cada evaluación realizada (Cardozo, mimeo).

En el caso del GDF, las cuatro evaluaciones coordinadas por Evalúa DF en 2010 y 2011 varían en sus alcances y temas, pero todas contaron con TR específicos. Los equipos evaluadores fueron interdisciplinarios, la mayoría de buen nivel académico, adecuados al tema, pero solo algunos contaban con experiencia previa en trabajos de evaluación. Los montos pagados fueron de USD 27 532 para los estudios de gabinete en 2010 y USD 84 541 para el que incluyó trabajo de campo en 2011.

Los trabajos contratados evaluaron políticas y servicios públicos y utilizaron fundamentalmente entrevistas y análisis de documentos como mé-

todos de investigación, enriquecidos con encuestas, observación *in situ*, grupos focales y análisis estadístico en el caso de la evaluación de impacto.

Los informes fueron relativamente extensos (entre 186 y 395 páginas) y, en general, se presentan adecuadamente estructurados y redactados, aunque subsisten algunos aspectos confusos; están disponibles en la página electrónica de Evalúa DF, donde se incluye además la respuesta institucional ante la evaluación y sus recomendaciones.

Se incluyen trabajos aceptables y otros de buena calidad, de difícil comparación por sus diferencias de complejidad técnica. Como consecuencia de lo anterior, las propuestas para los problemas organizacionales identificados son también variadas e incluyen: a) exigir mayor experiencia en evaluación a los equipos contratados; b) identificar claramente en todos los casos la relación entre datos, análisis, conclusiones, valoraciones y recomendaciones; c) especificar la instancia responsable y el plazo para cumplir las recomendaciones; d) acompañar siempre el informe final de un resumen ejecutivo; y e) dar seguimiento continuo a las recomendaciones aceptadas y proceder a iniciar las controversias correspondientes en los casos de rechazo (Cardozo, mimeo).

De la comparación de los trabajos de Evalúa DF con los revisados a nivel federal destaca una mayor presencia de convocatorias abiertas, acompañada de TR específicos para cada caso y el preestablecimiento del monto de los contratos, así como evaluaciones dirigidas a políticas públicas y servicios (no solamente programas). Los trabajos también son más amplios e intentan profundizar en la explicación de los hallazgos reportados. Solo en algunos pocos casos federales este resultado es explicable por su menor disponibilidad de recursos.

6. Utilización de resultados

No basta solo con tener evaluaciones y que estas sean de calidad, se necesita que se utilicen, especialmente para mejorar el diseño y el desempeño de los programas (utilización instrumental), por lo que las dependencias y entidades de la Administración Pública Federal se comprometieron a cumplir con un conjunto de Aspectos Susceptibles de Mejora (ASM) para aumentar la calidad de sus programas en función de las recomendaciones recibidas. Sin embargo, el propio Programa para un Gobierno Cercano y Moderno 2013-2018 (Diario Oficial de la Federación, 2013) plantea que en muchas ocasiones los ASM son vistos solamente como un procedi-

miento más a realizar en lugar de considerarlos una oportunidad de mejora para el programa. Además, menciona la importancia del proceso de evaluación para rendir cuentas a la ciudadanía, ofreciendo información de acceso público para la promoción de la participación ciudadana y propone un conjunto de estrategias y líneas de acción tendientes a articular, en forma transversal y multidisciplinaria, las diferentes acciones de las dependencias y entidades en la materia.

A título de ejemplo en 2012-2013, el porcentaje de recomendaciones federales acordadas con los operadores para ser transformadas en ASM fue de 45 %. Además, desde 2007 en que se emitieron 930 ASM, estos muestran una clara tendencia a la disminución, con un aumento pasajero en 2011-2012, en que se efectuaron ECR. En consistencia con la reducción anterior de ASM y un número importante de cancelaciones (83 ASM en 2013-2014), su nivel de cumplimiento ha ido en aumento hasta alcanzar el 82 %, con una disminución en 2012-2013 que coincide con la aplicación de las recomendaciones surgidas de la ECR. Por otra parte, los ASM se concentran cada vez más en las correcciones de programas, mientras las reorientaciones vinculadas al rediseño se mantienen en solo un 22 % (ver Cuadro N° 6).

Cuadro N° 6. Seguimiento de ASM desde 2010
(cumplidos a septiembre de 2014)

Avances	Informe 2010-2011	Informe 2011-2012	Informe 2012-2013	Informe 2013-2014
Número de ASM formulados	413	471	641	260
Promedio general de avance (%)	69.4	73.9	62.6	81.9
Impactos de los ASM (%)	Adiciones 36 Correcciones 22 Modificaciones 18 Reorientaciones 22 Suspensión 2	Adiciones 0 Correcciones 71 Modificaciones 12 Reorientaciones 15 Suspensión 1	Adiciones 4 Correcciones 64 Modificaciones 4 Reorientaciones 28 Suspensión 0	Adiciones 0 Correcciones 72 Modificaciones 4 Reorientaciones 22 Suspensión 0

Fuente: Informes de Seguimiento de ASM de 2010-2011 a 2013-2014, en Coneval (s/f)

En síntesis, a pesar del carácter no vinculatorio de las decisiones de Coneval y de enfrentar fuertes resistencias burocráticas, la utilización instrumental de resultados ha alcanzado un nivel relativamente importante de cumplimiento cuantitativo, quedando pendiente la evaluación del impacto que efectivamente han tenido en las políticas y programas y fundamentalmente, en sus beneficiarios. Sin embargo, es importante resaltar que esto es fruto de un proceso en que Coneval acuerda los ASM con los operadores (ellos toman parte en la decisión de lo que pueden y están dispuestos a hacer), que el nivel de transformación de las recomendaciones en ASM es muy bajo y que estos solo provocan reorientaciones de programas en un porcentaje muy reducido.

En relación con la utilización presupuestal, con base tanto en los casos mexicanos analizados por Cardozo (mimeo) como en el estudio efectuado por Székely (2013) que recoge la opinión de una muestra de funcionarios, no se percibe incidencia de las evaluaciones en la asignación de recursos.

Por otro lado, en relación con la utilización instrumental en el GDF, el número de recomendaciones vinculatorias enviadas por Evalúa DF fue siempre mayor que las sugeridas en los informes de evaluación debido a que este Consejo las redacta complementando las señaladas en el informe. Sin embargo, las dependencias evaluadas tienen la posibilidad de manifestar su desacuerdo, en cuyo caso Evalúa DF puede controvertirlas.

La aceptación de las recomendaciones ligadas a la operación de los programas o la implementación de las políticas (cambios en los procedimientos en curso) era mayor que la ligada a su reformulación (modificación del diseño mismo de la política) que, en muchos casos, implica mayor disponibilidad de recursos. El Cuadro N° 7 muestra la evolución de las recomendaciones 2009-2011 y el N° 8 el nivel de cumplimiento de las recomendaciones 2009 en junio de 2011.

Cuadro N° 7. **Recomendaciones 2009-2011**

Año	Recomendaciones emitidas	Recomendaciones aceptadas	Recomendaciones reconsideradas	Preparadas para controversia
2009	124	105 (84.6 %)	5	14
2010	35	32 (91.4 %)	3	0
2011	77	69 (89.6 %)	0	8

Fuente: Primer Informe Semestral del Estado de las Recomendaciones Aceptadas, en Evalúa DF (2011).

En 2009 las 124 recomendaciones fueron dirigidas a 12 programas y dependencias, se distribuyeron aproximadamente en la misma proporción en materia de diseño (49.6 %) que de operación (50.4 %); sin embargo, de las 14 no aceptadas (11.3 %), nueve se concentraron en diseño y ocho de ellas implicaban una fuerte repercusión presupuestal. Todas las controvertidas correspondían a dos de los programas evaluados.

Las cuatro políticas y servicios evaluados en 2010 y 2011 arrojaron un 47 % de aceptación plena y un 18 % de aceptación parcial, con lo que se estima un 56 % de aceptación global, bastante inferior a los porcentajes reportados en años previos. En 2010, las tres recomendaciones no aceptadas correspondieron al mismo programa; sin embargo, en los meses siguientes se modificó la ley correspondiente recuperando la esencia de las mismas, razón por la que ya no fueron controvertidas por Evalúa DF. En este caso se hizo la primera experiencia de dirigir recomendaciones directamente a la Secretaría de Finanzas cuando existían importantes repercusiones en las necesidades financieras.

A partir de 2011 las recomendaciones dejaron de dirigirse a programas para hacerlo a políticas, al igual que las evaluaciones realizadas, lo que complicó el seguimiento porque las recomendaciones ya no se dirigían a un solo órgano (responsable en el caso de los programas) sino a las varias instancias coordinadas para poner en marcha una política. Se volvieron a enviar recomendaciones a la Secretaría de Finanzas, debido a que los órganos operativos no aceptaban las recomendaciones de fuerte impacto presupuestal por temor de que dicha Secretaría no les autorizara el gasto correspondiente, pero esta empezó también a rechazarlas por considerar que no le competía emitir opinión sobre las recomendaciones de carácter operativo, aunque ellas tuvieran fuerte impacto presupuestal.

Cuadro N° 8. Estado de las recomendaciones 2009 al 30 de junio de 2011

Recomendaciones	Número	Porcentaje
Totalmente cumplidas	65	63
Parcialmente cumplidas	20	19
No cumplidas	18	17
Aún vigentes	2	1
Total	105	100

Fuente: Primer Informe Semestral del Estado de las Recomendaciones Aceptadas, en Evalúa DF (2011).

En general, el cumplimiento total o parcial de la primera generación de recomendaciones (2009) a las evaluaciones iniciadas en 2008 fue bastante elevado al 30 de junio de 2011³ (73 %, suponiendo que el cumplimiento parcial fuera del 50 %), aunque existía un importante grado de retraso en su implementación y en la entrega de la información correspondiente. En consecuencia, tampoco el carácter vinculatorio de Evalúa DF le ha permitido, hasta la fecha, tener la contundencia esperada en materia de utilización instrumental de las evaluaciones.

Por otra parte, Evalúa DF debe entregar copia de los informes de evaluación a la Comisión de Desarrollo Social de la Asamblea Legislativa del DF y esta los ha utilizado en diversos casos, mas no con fines presupuestales. Más bien han sido las propias instancias evaluadas quienes argumentan falta de recursos para negarse a llevar a cabo las recomendaciones, por temor de que la Secretaría de Finanzas no les apruebe el incremento presupuestal requerido. En consecuencia, tampoco el proceso de aceptación y cumplimiento de las recomendaciones muestra impactos sustantivos para mejorar la distribución de los recursos presupuestales en ninguno de los dos casos considerados, federal ni del GDF.

Pero, además de asegurar la puesta en práctica de las buenas recomendaciones surgidas de un proceso de evaluación, es necesario verificar el impacto que ellas están teniendo en el logro de los objetivos planeados para resolver problemas sociales. Tal vez sea muy pronto para que esta actividad se encare, pero a la fecha ninguno de los dos Consejos de evaluación considerados en el trabajo ha avanzado en su análisis y valoración.

Conclusiones

Desarrollar e internalizar una cultura de la evaluación en diferentes grupos sociales (funcionarios, diputados, investigadores, ONGs) y la ciudadanía en general constituye la máxima aspiración de quienes estamos convencidos de su necesidad y conveniencia, para que esta sea voluntariamente aplicada, desarrollada en forma colaborativa, sus productos sean de alta calidad y sus resultados ampliamente utilizados. Pero lograrlo requiere tiempos muy prolongados como ha sucedido en EUA y, en menor grado, en Reino Unido y Canadá, por lo que los países latinoamericanos

³ Posteriormente no se ha publicado información, razón por la que no ha sido posible cubrir hasta el año 2014, como se hizo a nivel federal.

que han iniciado recientemente su institucionalización formal (Colombia, Chile, México, Brasil, Perú, Argentina, etc.) tienen aún un largo camino por recorrer.

Para avanzar hacia la meta planteada es fundamental contar con espacios de profesionalización y reflexión sobre el tema: programas de formación y capacitación, asociaciones de profesionales que organicen eventos y discutan temas relevantes y publicación de libros y revistas especializadas que difundan nuevas ideas, métodos y experiencias. En este sentido, los países de América Latina se caracterizan por sus grandes limitaciones.

La exposición realizada muestra a México como un país que inició un poco tarde sus procesos de evaluación pero ha crecido aceleradamente después de la creación de Coneval en 2006, y atraviesa un momento de incertidumbre relacionada con las posibles consecuencias de los cambios legales que estarán afectando el funcionamiento de dicho Consejo, con el posible riesgo de pérdida de independencia, no legal pero sí práctica.

Junto con este notable aumento del número de evaluaciones realizadas, la tendencia ha sido hacia la estandarización, la simplificación y el abaratamiento de los trabajos, lo que lleva a cuestionarse sobre su calidad y el riesgo de que sean considerados un procedimiento más, sin mayores consecuencias posteriores. En el caso del GDF también es muy importante cuidar que esta experiencia reciente y novedosa mantenga su calidad, nivel de independencia y aumente su utilización de resultados.

En la revisión de todos los ciclos mexicanos de evaluación de programas federales 2008-2014 aparecen siempre trabajos de buena y muy buena calidad (especialmente las ECR, entre las más frecuentes), pero también algunos que carecen de solidez metodológica y de aportaciones (EED, también entre las frecuentes). En sus procesos suelen detectarse problemas que Evalúa DF ha tratado de evitar como: falta de TR o disponibilidad solo de términos genéricos; adjudicaciones directas de contratos; equipos de evaluación cuya formación, experiencia y composición disciplinaria no siempre es la adecuada; costos muy diversos y no justificados; ausencia de trabajo de campo y de recuperación de percepciones de los beneficiarios; tamaño irrisorio de algunos informes, entre otros.

La situación federal anterior contrasta con la vivida en 2001-2006 en que, a pesar de la inexperiencia se hacía un gran esfuerzo por llevar a cabo evaluaciones integrales, con trabajo de campo e informes más amplios y también con la evaluación realizada en el GDF en 2008-2014 que cuenta siem-

pre con convocatorias abiertas, TR específicos, montos preestablecidos e iguales para trabajos similares, e informes más amplios y explicativos, aunque también exhibe trabajos de calidad reducida. En los dos casos mencionados, esto ha sido posible debido a que se ha realizado un pequeño número de evaluaciones, lo que no sucede en la práctica federal actual y que podría resolverse mediante la creación de diversos órganos independientes de evaluación especializados por secretarías (e.g. Instituto Nacional de Evaluación Educativa), entidades (e.g. Evalúa DF) y municipios.

En síntesis, la situación federal actual muestra una tendencia a cumplir con el requisito formal de realizar una evaluación anual de un número muy elevado de programas, cuya calidad no está garantizada y, en consecuencia, tampoco está clara la conveniencia de utilizar sus recomendaciones. Esto se acompaña también de una ampliación de las atribuciones de Coneval (asesoría a gobiernos subnacionales, evaluación de consejos estatales), lo que contribuye a privilegiar la cantidad sobre la profundidad. En cambio, Evalúa DF puede elegir un número reducido de programas a evaluar y aplicar metaevaluaciones para mejorar la calidad de los trabajos, pero tampoco ha logrado la incidencia esperada en cuanto al uso de sus resultados. Constituye una buena alternativa pero aún incipiente y vulnerable si no se la apoya consistentemente.

Al respecto, se encontró que el porcentaje de cumplimiento de las recomendaciones comprometidas a nivel instrumental federal es relativamente alto, pero está limitado por la calidad de las sugerencias realizadas por los equipos evaluadores y el bajo porcentaje de ellas asumido como compromiso, lo que lleva a otra discusión: la conveniencia de que las recomendaciones emitidas por los Consejos de evaluación tengan carácter vinculatorio. Sin embargo, el sistema implantado por Evalúa DF ya lo ha puesto en práctica con resultados no tan positivos como se esperaba debido a la poca efectividad de las sanciones previstas.

Finalmente, la teoría y las experiencias mundiales muestran un sinnúmero de factores que pueden explicar el limitado nivel de utilización alcanzado, tantas veces señalado en la metaevaluaciones de distintos países. En estos trabajos, la calidad de las evaluaciones no se menciona con un peso significativo como determinante de la utilización; sin embargo, tampoco interesa promover la utilización al margen de ella, por lo que se requieren trabajos de metaevaluación. Por otra parte, aunque todos los factores señalados en la literatura son útiles por su contribución a la explicación *ex post* de la utilización lograda, nuestro interés se centra en aquellos que pueden ser modificados en el corto plazo por el propio sistema de evaluación, de manera de incrementar las posibilidades de que las buenas evaluaciones sean tomadas en cuenta. En

consecuencia, se seleccionó un subconjunto de aspectos que podrían contribuir a incrementar dicha utilización en el contexto mexicano como: amplia participación de funcionarios desde la planeación de los trabajos, selección de evaluadores en concursos abiertos, levantamiento de percepciones de beneficiarios y organizaciones sociales, presentación oportuna, accesible y clara del informe, vinculación con la utilización presupuestal, incentivos monetarios y simbólicos, colectivos e individuales, así como seguimiento continuo del uso comprometido de las recomendaciones (Cardozo, mimeo).

En síntesis, es necesario seguir trabajando para desarrollar una cultura de la evaluación apoyada con más programas universitarios, publicaciones y eventos pero, mientras esto no se alcanza, hay que mejorar por medios alternativos tanto la calidad de los trabajos (en riesgo cuando hay un crecimiento muy importante de la cantidad) como su posterior utilización, para que los esfuerzos realizados cobren el sentido que justifica la existencia del trabajo evaluativo en nuestro contexto.



Bibliografía

- BANCO MUNDIAL (s/f) “México: Análisis de los Lineamientos generales para la evaluación de los programas federales de la administración pública federal”, disponible en: <http://www.coneval.gob.mx/rw/resource/coneval/quienes/2937.pdf>, consultada el 14 de mayo de 2013.
- CARDOZO, M. (2006) *La evaluación de políticas y programas públicos. El caso de los programas de desarrollo social en México*. Colección Conocer para Decidir, México: Cámara de Diputados-M. A. Porrúa.
- . (2009) *Evaluación y metaevaluación en los programas mexicanos de desarrollo social*. Colección Teoría y Análisis, México: UAM-X.
- . (2011) “Organización y métodos alternativos en la evaluación de políticas y programas sociales. La experiencia consolidada y la emergente en el caso de México”, en *Revista Gestión y Análisis de Políticas Públicas* N° 5, 43-59.
- . (2012a). *Evaluación y metaevaluación en las políticas y programas públicos. El estado del arte*. México: UAM-X.
- . (2012b) *Sistematización de la experiencia de evaluación de Evalúa DF en 2008. Metaevaluación y autoanálisis*. México: UAM-X- Evalúa DF.
- . (mimeo) *Institucionalización de procesos de evaluación. Calidad y utilización de sus resultados*, México: UAM-X.
- CARDOZO, M. y MUNDO, A. (2012) *Guía de orientación para la evaluación de políticas y programas de desarrollo social*. México: Evalúa DF.
- CEJUDO, G. (2014) “Implicaciones de la autonomía del Coneval”, disponible en: <http://rendiciondecuentas.org.mx/implicaciones-de-la-autonomia-constitucional-de-coneval/>, consultada el 03 de septiembre de 2014.

- CLEAR (s/f) <http://www.clear-la.cide.edu/node/6>, consulta efectuada el 12 e diciembre de 2014.
- CONEVAL (2013) “Modelo de términos de referencia para la evaluación específica de desempeño 2012-2013”, disponible en: [www.google.com/url?q=http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/NORMATIVIDAD/TdR_EED_2012/EED_TdR_2012_2013.pdf&sa=U&ei=xd9kVMj6OZWWhugS6pYDoAw&ved=0CAkQFjADOB4&client=internal-uds-cse&usg=AFQjCNFWz_K8tp0j-3hwDib-7B3AQ4R3nA](http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/NORMATIVIDAD/TdR_EED_2012/EED_TdR_2012_2013.pdf&sa=U&ei=xd9kVMj6OZWWhugS6pYDoAw&ved=0CAkQFjADOB4&client=internal-uds-cse&usg=AFQjCNFWz_K8tp0j-3hwDib-7B3AQ4R3nA), consultada el 12 de noviembre de 2014.
- CONEVAL (s/f) <http://www.coneval.gob.mx>, múltiples consultas, última realizada el 17 de febrero de 2015.
- CONEVAL-SHCP-SFP (2007) *Lineamientos generales para la evaluación de los programas federales de la administración pública federal*, disponible en: http://www.coneval.gob.mx/rw/resource/coneval/eval_mon/361.pdf, consultada el 03 de septiembre de 2014.
- EVALÚA DF (2011) *Primer informe semestral del estado de las recomendaciones aceptadas*, mimeo.
- . (s/f) <http://www.evalua.DF.gob.mx>, múltiples consultas, última realizada el 15 de enero de 2015.
- GARCÍA y GARCÍA (2010). *La gestión para resultados en el desarrollo. Avances y desafíos para América Latina y el Caribe*. BID.
- LEY GENERAL DE DESARROLLO SOCIAL. *Diario Oficial de la Federación*, 20 de enero de 2004 (última reforma 07 de noviembre de 2013).
- LEY DE DESARROLLO SOCIAL PARA EL DISTRITO FEDERAL. *Gaceta Oficial del Distrito Federal*, 23 de mayo de 2000.
- PROGRAMA PARA UN GOBIERNO CERCANO Y MODERNO 2013-2018. *Diario Oficial de la Federación*, 30 de agosto de 2013.
- ROJAS, F. (2013). “Evaluación de desempeño organizacional del Consejo Nacional de Evaluación de la Política de Desarrollo Social”, disponible en: http://www.coneval.gob.mx/Informes/Coordinacion/Evaluaciones_auditorias_externas/EDO_Resumen_Ejecutivo.pdf, consultada el 17 de noviembre de 2014.
- SHCP (2012). Evaluación del sistema de presupuesto por resultados 2012, disponible en: http://www.siap.gob.mx/wp-content/pdf/programas_plan_desarrollo/ayuda.pdf, consultada el 14 de octubre de 2014.
- SZÉKELY, M. (2013). “Evaluación externa de los resultados del Consejo Nacional de Evaluación de la Política de Desarrollo Social”, disponible en: http://www.coneval.gob.mx/Informes/Coordinacion/Evaluaciones_auditorias_externas/Informe_EER.pdf, consultada el 17 de noviembre de 2014.

Fecha de recepción: 02/03/2015

Fecha de aceptación: 30/04/2015