

Institucionalización de procesos de **evaluación**

Calidad y utilización de sus resultados

Myriam Cardozo Brum
coordinadora

Colección **Teoría y Análisis**



UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD XOCOMILCO División de Ciencias Sociales y Humanidades



Centro de Estudios
CSOP
Instituto de la Opinión Pública

INSTITUCIONALIZACIÓN DE PROCESOS DE EVALUACIÓN
CALIDAD Y UTILIZACIÓN DE SUS RESULTADOS

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: 29 de noviembre de 2015

D.R. © Universidad Autónoma Metropolitana

UAM-Xochimilco

Calzada del Hueso 1100

Col. Villa Quietud, Coyoacán

C.P. 04960 México, DF.

Sección de Publicaciones,

División de Ciencias Sociales y Humanidades

Edificio A, tercer piso. Teléfono 5483 7060

[pubcsh@correo.xoc.uam.mx] [<http://dcshpublicaciones.xoc.uam.mx>]

ISBN: 978-607-28-0546-0

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

Centro de Estudios Sociales y de Opinión Pública

Cámara de Diputados / LXIII Legislatura

Av. Congreso de la Unión 66

Edificio I, Primer Piso

Col. El Parque

México, D.F.

Tel. 5036-0000 ext. 55237

[cesop@congreso.gob.mx]

[<http://diputados.gob.mx/cesop>]

Impreso y hecho en México / Printed and made in Mexico

Institucionalización de procesos de evaluación

Calidad y utilización de sus resultados

Myriam Cardozo Brum
(coordinadora)

Talina Olvera Mejía / Angélica Rosas Huerta
Esthela Sotelo Núñez / Laura Rodríguez Cardozo
Osmar Medina Urzúa / Myriam Cardozo Brum





UNIVERSIDAD AUTÓNOMA METROPOLITANA

Rector general, Salvador Vega y León

Secretario general, Norberto Manjarrez Álvarez

UNIVERSIDAD AUTÓNOMA METROPOLITANA-XOCHIMILCO

Rectora de Unidad, Patricia E. Alfaro Moctezuma

Secretario de Unidad, Joaquín Jiménez Mercado

DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES

Director, Carlos Alfonso Hernández Gómez

Secretario académico, Alfonso León Pérez

Jefa del departamento de Política y Cultura, Alejandra Toscana Aparicio

Jefe de la sección de publicaciones, Miguel Ángel Hinojosa Carranza

CONSEJO EDITORIAL

Aleida Azamar Alonso / Gabriela Dutrénit Bielous

Diego Lizarazo Arias / Graciela Y. Pérez-Gavilán Rojas

José Alberto Sánchez Martínez

Asesores del Consejo Editorial: Luciano Concheiro Bórquez

Verónica Gil Montes / Miguel Ángel Hinojosa Carranza

COMITÉ EDITORIAL DEPARTAMENTAL

Pablo Mejía Montes de Oca (presidente)

Clara Martha Adalid Urdanivia / Juan José Carrillo Nieto

Marco Antonio Molina Zamora / Esthela Sotelo Núñez

Luis Miguel Valdivia Santamaría / Ricardo Yoclevzky Retamal

Asistencia editorial: Varinia Cortés Rodríguez

Diseño de portada: Irais Hernández Güereca

Índice

Presentación	11
Introducción general	13

PRIMERA PARTE

Institucionalización y profesionalización en evaluación

Introducción	19
--------------------	----

Países de Europa y de Norteamérica

I. Estados Unidos de América	23
II. Reino Unido	39
III. Francia	49
IV. Canadá	67

América Latina

V. Colombia	81
VI. Costa Rica	91
VII. Chile	99

México

VIII. Institucionalización y profesionalización de la evaluación	109
IX. Desarrollo social	125

X. Educación	131
XI. Medio ambiente	143
XII. Salud	159
XIII. Desarrollo rural	165
Conclusiones generales de la primera parte	173

SEGUNDA PARTE

Metaevaluación: calidad de las evaluaciones aplicadas en México 2007-2014

Introducción	183
XIV. Evaluación y metaevaluación	185
Normatividad y tipos de evaluación	186
Metaevaluación: selección de evaluaciones a analizar	196
XV. Metaevaluaciones en el área de desarrollo social federal	203
Procesos del Programa Fomento y Desarrollo de Culturas Indígenas 2008-2009	203
Impacto de Oportunidades 2008-2009	208
Integral de desempeño de Ayuda Alimentaria y Generación de Capacidades 2010-2011	216
Impacto de Estancias Infantiles para apoyar a Madres Trabajadoras 2011	220
Estratégica de Protección Social en México 2013	226
Conclusiones y recomendaciones parciales	232
XVI. Metaevaluaciones en materia educativa federal	235
Específica de desempeño de Escuelas de Tiempo Completo 2009-2010	235
Impacto de Escuelas de Calidad 2012	240
Específica de Escuelas Dignas 2013	246
Diagnóstico de Escuelas de Excelencia para abatir el Rezago Educativo 2014	249
Conclusiones y recomendaciones parciales	254
XVII. Metaevaluaciones en el campo ambiental federal	257
Indicadores de mediano plazo del Programa Nacional de Áreas Naturales Protegidas 2009	257

Diseño del Programa de Manejo de Áreas Naturales Protegidas 2011-2012	266
Diseño del Programa Estabilización de Cuencas y Acuíferos 2013	270
Conclusiones y recomendaciones parciales	280
XVIII. Metaevaluaciones en materia de salud federal	283
Consistencia y resultados del Sistema de Protección Social en Salud. Seguro Popular de Salud 2007	283
Complementaria del Sistema Integral de Calidad en Salud 2008-2009	287
Conclusiones y recomendaciones parciales	292
XIX. Metaevaluaciones de desarrollo rural federal	295
Consistencia y resultados del Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural 2011-2012	295
Consistencia y resultados del Programa de Apoyo al Ingreso Agropecuario: Procampo para Vivir Mejor 2011-2012	303
Conclusiones y recomendaciones parciales	309
XX. Metaevaluación de evaluaciones del Gobierno del Distrito Federal	311
Evaluaciones en áreas de política social, transporte y salud 2010-2011	311
Desempeño del Programa de Cambio Climático de la Ciudad de México 2008-2012	337
Conclusiones y recomendaciones parciales	348
Conclusiones y recomendaciones generales de la segunda parte	351

TERCERA PARTE
Utilización de las evaluaciones

Introducción	361
XXI. Revisión de la literatura	363
Estados Unidos y Canadá durante el siglo XX	364
Proliferación mundial de estudios en el siglo XXI	370
España y América Latina	374
Síntesis	378

XXII. Análisis de experiencias	383
Países exitosos	383
México	385
Conclusiones de la tercera parte	437
Síntesis general de conclusiones	445
Referencias	449
Siglas y acrónimos	483
Índice de cuadros y gráficas	491

Presentación

La Constitución Política de la República Mexicana es una guía para garantizar los derechos de los ciudadanos, entre ellos el derecho a que se les rinda cuentas acerca del buen uso de los recursos provenientes de la cooperación social. Dado que la Constitución ha sido constantemente reformada por el Constituyente Permanente y que el proceso para reformarla requiere de la aprobación de dos terceras partes del Congreso y de la mayoría de los congresos de los estados, este sistema de mayoría calificada permite afirmar que sus reformas son discutidas y aceptadas por los representantes de buena parte de los ciudadanos del país.

Ahora bien, estas reformas nos permiten trazar la historia de las preocupaciones sociales de los ciudadanos. Una de las preocupaciones clave desde la Constitución de 1824 que estableció la República democrática, fue que el Congreso revisara la cuenta pública ejercida por el Ejecutivo. En un principio simple de división de poderes, los representantes de los ciudadanos aprueban los impuestos y el destino de éstos y supervisan que su gasto se ejerza adecuadamente.

Claramente, el cambio social ha hecho que la rendición de cuentas del Ejecutivo al Legislativo sea cada vez más compleja, requiriendo instituciones nuevas que supervisen y reporten el gasto. En este libro se consigna la historia, tanto en México como en otros países, de un esfuerzo por hacer de la rendición de cuentas un proceso que redunde en un mejor gasto en beneficio de los ciudadanos. Sin embargo, una de las conclusiones centrales del libro es que las nuevas instituciones evaluadoras requieren de un complejo entramado administrativo y político que las relacione con las viejas instituciones administradoras de los recursos. Es evidente que los cambios sociales han sido

tan rápidos, que dar seguimiento preciso a los efectos del gasto y responder con cambios en las acciones de gobierno de acuerdo con las evaluaciones de estos efectos, aún está muy lejos de ser la norma.

Este libro contribuye a un debate constante en el seno del Poder Legislativo: se trata de los alcances que debe tener la revisión de cuentas sobre el otro poder electo, el Ejecutivo. Es evidente que ahí donde haya malversación ésta se persiga; pero cómo determinar si el gasto a corto plazo no tendrá efectos benéficos en el largo plazo, o si es prerrogativa del Ejecutivo ejercer con cierta discrecionalidad con fines intangibles, como la paz social. De ahí que el Centro de Estudios Sociales y de Opinión Pública de la Cámara de Diputados se haya interesado en coeditar este libro y así contribuir al debate y a la discusión permanente sobre la mejora de las evaluaciones del gasto, sus alcances, sus costos y sus efectos sobre la política pública.

Introducción general

Con motivo del Año Internacional de la Evaluación (2015), promovido por The International Evaluation Partnership Initiative (EvalPartners)¹ y asumido por organismos internacionales, gobiernos nacionales, fundaciones privadas y sociedades de evaluación² con la finalidad de realizar múltiples actividades tendientes a poner a la evaluación en el centro del debate público, y considerando la importancia adquirida por los procesos de evaluación de las políticas y programas gubernamentales, se ha querido contribuir con un libro dedicado a actualizar tanto las discusiones teóricas como las experiencias prácticas, especialmente en el contexto mexicano.

La evaluación permite conocer el nivel de desempeño de políticas y programas (resultados e impactos), explicarlo tomando en consideración las características de su diseño e instrumentación pero también del contexto político, social y

¹ EvalPartners es una iniciativa del Fondo de las Naciones Unidas para la Infancia (UNICEF), la Organización Internacional de Cooperación en Evaluación (IOCE) y otros organismos para impulsar la capacidad de evaluación de las Organizaciones de la Sociedad Civil (OSC).

² Durante la III Conferencia Internacional de Capacidades Nacionales en Evaluación realizada en San Pablo, Brasil, en septiembre-octubre de 2013, los participantes asumieron como suya la propuesta de EvalPartners, en especial la Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (Relac); así, 2015 fue considerado idóneo para ser el Año Internacional de la Evaluación porque durante él serán analizados los logros alcanzados por los Objetivos del Milenio y reemplazados por otros nuevos.

económico concreto en que se inserta, y valorarlo a la luz de un conjunto de criterios como la pertinencia, la congruencia, la eficacia, la eficiencia, la calidad, la equidad, entre otros.

La información que la evaluación genera cubre diferentes necesidades: difusión, para cumplir con la obligación del gobierno de rendir cuentas a la ciudadanía y que ésta cuente con la información que le permita incorporarse a los procesos participativos; instrumental, para mejorar la gestión de las políticas y programas evaluados; presupuestal, para asignar los recursos en función de los resultados obtenidos; conceptual, para mejorar la comprensión general de los principales componentes y relaciones; y persuasión, para aumentar la contundencia de un argumento con la intención de convencer a otros.

La evaluación requiere de: *a*) un proceso de institucionalización que facilite el desarrollo de una investigación aplicada e interdisciplinaria por parte de un equipo con formación y experiencia tanto en sus métodos como en torno al problema que es objeto de la acción pública a analizar; *b*) mecanismos que garanticen la calidad de los trabajos y la factibilidad de sus recomendaciones; y *c*) la utilización de sus resultados para satisfacer las necesidades antes planteadas. Estos tres aspectos constituyen los temas tratados en cada una de las partes del libro.

La primera parte realiza un recorrido por diversos países para conocer el nivel de institucionalización alcanzado por la evaluación en sus contextos nacionales, mostrando avances importantes en América (Estados Unidos y Canadá) y Europa (Reino Unido y Francia) e impulsos relativamente recientes en América Latina (Colombia, Costa Rica, Chile y México). En cada caso se presenta un breve resumen de lo acontecido durante el siglo XX para luego revisar más detalladamente los desarrollos recientes. En el caso particular de México, el estudio se hace con mayor profundidad, analizando tanto la situación a nivel federal como en el caso de la capital del país y, a su interior, desglosándola por sectores: desarrollo social, educación, salud, medio ambiente, transporte y desarrollo rural.

Una de las reflexiones que acompañan este primer tramo, aunado a la institucionalización a nivel gubernamental, es que se requiere de un trabajo paralelo que contribuya a la profesionalización de los procesos de evaluación a través de la disposición de programas universitarios, publicación de libros y revistas especializadas y creación de asociaciones de profesionales. Este trabajo, desarrollado a lo largo de 13 capítulos, permite sobre todo una medición

y comparación bastante cuantitativa del nivel de avance logrado (montos de recursos, cantidad de evaluaciones realizadas, número de programas de estudio, de publicaciones, de sociedades profesionales, entre otros) pero, falta acompañarlo del estudio de la calidad de lo realizado, lo que nos lleva al contenido de la segunda parte.

Ésta se conforma de ocho capítulos dedicados a realizar la metaevaluación de: *a*) una muestra³ de 16 evaluaciones de programas llevadas a cabo a nivel federal en 2007-2014, elegidas con base en diversos criterios; (sectores diversos de políticas y de programas,⁴ tipos diferentes de evaluación⁵ y distribución anual) y *b*) la totalidad de las cinco evaluaciones de políticas, programas y servicios efectuadas en 2010-2011⁶ a nivel de la capital, incluyendo sectores prioritarios del gobierno.⁷ En el conjunto de estos 21 trabajos se identificó un número importante de evaluaciones que no tienen la calidad necesaria, por lo que no resulta conveniente que sus recomendaciones se apliquen, y otro conjunto de buena calidad del que se investiga si ha sido efectivamente utilizado para mejorar las políticas y los programas públicos.

Dicha utilización constituye el tema analizado en la tercera parte, mediante la revisión de trabajos teóricos y aplicados tanto a nivel mundial como de los diferentes sistemas y sectores mexicanos. El primero de sus capítulos muestra una enorme variedad de posibles factores explicativos de la utilización de resultados de evaluación que incluye entre los más mencionados: clima político, normatividad, relaciones intergubernamentales, cultura de evaluación, credibilidad de los evaluadores, claridad y difusión del informe, sistemas de incentivos, etcétera. El segundo capítulo se dedica a identificar el nivel de utilización instrumental, presupuestal y conceptual en el caso mexicano. A nivel federal la información

³ Dado el tamaño del universo de evaluaciones realizadas en este periodo fue imposible trabajar con una muestra representativa de las mismas.

⁴ Desarrollo social, educación, medio ambiente, salud y desarrollo rural.

⁵ De diagnóstico, diseño, específicas de desempeño, consistencia y resultados, procesos, indicadores, integral de desempeño, complementaria, estratégica y de impacto.

⁶ Si bien éstas empezaron a realizarse en 2008, las previas a 2010 ya habían sido sometidas a metaevaluación y las posteriores a 2011 se tienen planeadas recién para el año en curso (2015).

⁷ Desarrollo social, transporte, salud y medio ambiente.

disponible no permite un seguimiento por evaluación en función de la calidad del trabajo, sino sólo por grupos sectoriales. Se señala un cumplimiento cuantitativo importante de las recomendaciones asumidas por los operadores, pero bastante bajo de las sugerencias de los evaluadores que los funcionarios operativos de los programas acuerdan llevar a cabo, así como una concentración en ajustes menores en detrimento de modificaciones de fondo.

Esta excesiva decantación de las sugerencias lleva nuevamente a tratar de identificar en la literatura los factores que pueden contribuir a la mayor utilización. Pero, debido a que muchos de los factores señalados no pueden ser modificados, especialmente en el corto plazo, por los actuales sistemas de evaluación, se seleccionó un subconjunto de ellos que sí pueden modificarse para facilitar la utilización inmediata de los resultados de la evaluación de buena calidad en México. Dichos factores incluyen: 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.

PRIMERA PARTE
Institucionalización y profesionalización en evaluación

Introducción

Desde inicios del siglo XX es posible identificar esfuerzos pioneros en el desarrollo de trabajos evaluativos, como en el caso del Reino Unido en torno a los servicios públicos o de los Estados Unidos de América en materia de servicios educativos. En las siguientes décadas se fueron agregando nuevos temas para incluir una variedad de programas sociales.

Sin embargo, el auge de esta actividad se incrementó notablemente desde la década de 1970 a partir de su reconocimiento paulatino por parte de los organismos que conforman la Organización de las Naciones Unidas (Organización de las Naciones Unidas para la Alimentación y la Agricultura, Fondo de las Naciones Unidas para la Infancia, etcétera), bancos multilaterales de fomento (Banco Mundial, Fondo Monetario Internacional, Banco Interamericano de Desarrollo), la Unión Europea (principalmente Dinamarca, Países Bajos, Francia, Alemania, Irlanda, etcétera) y países de otras regiones (Canadá, Costa Rica, Colombia).

La mayoría de las naciones ha intentado aplicar la evaluación a partir de la década de 1980, lo que fue reforzado con la segunda generación de reformas del Estado de la década de 1990 para tratar de recuperar la pérdida de legitimación de los gobiernos y debido al auge de la Nueva Gerencia Pública, a los logros alcanzados en materia de evaluación por los países de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y a la presión de los actores sociales. En paralelo, algunas regiones han formado profesionales universitarios especializados en evaluación. Países como Chile y México se incorporan a este proceso en los últimos años del siglo pasado.

Con objeto de mostrar los avances en el estado del arte en materia de evaluación, en esta parte del libro se seleccionó un grupo de países pioneros de Europa y Norteamérica y otro integrado por latinoamericanos y México en particular, para describir, analizar y valorar sus procesos de institucionalización formal de la evaluación y analizar su relación con la profesionalización alcanzada mediante programas académicos, publicación de trabajos especializados y debates en el marco de asociaciones de profesionales.

Países de Europa y de Norteamérica

Los países con mayor desarrollo han institucionalizado sus sistemas de evaluación a lo largo del siglo XX y han alcanzado distintos grados de consolidación. Resaltan los que integran la Unión Europea, Estados Unidos y Canadá, pero también otros más lejanos como Japón y Australia.

Ante la imposibilidad de actualizar el estado del arte en todos ellos, se han seleccionado algunos casos cuyas experiencias pueden aportar un aprendizaje de interés para el caso mexicano, y que por razones culturales y de proximidad geográfica permiten contar con mayor información y oportunidades de intercambio. Con base en estos criterios se han seleccionado dos países europeos (Reino Unido y Francia) y dos norteamericanos (Estados Unidos y Canadá), lo que además permite presentar tanto experiencias sajonas como latinas.

En cada uno de los casos mencionados se presenta un breve resumen de lo acontecido durante el siglo XX para luego revisar en detalle los desarrollos recientes. Se aborda la institucionalización y el nivel de consolidación alcanzado por la evaluación como una actividad profesional, refiriendo tanto a las acciones gubernamentales como al papel que desempeñan las asociaciones profesionales, las publicaciones especializadas más importantes y los programas académicos de nivel superior dedicados a la formación de profesionales de la evaluación.

CAPÍTULO I

Estados Unidos de América

*Talina Olvera Mejía**

La evaluación de programas y políticas públicas llevada a cabo en los Estados Unidos de América (EUA) se ha convertido en un referente, por lo que en este apartado se recoge parte de los esfuerzos que han logrado consolidar su institucionalización y profesionalización. Antes de presentar el panorama actual de la evaluación, se comienza por exponer brevemente lo ocurrido en materia de evaluación durante el siglo pasado.¹

Desde inicios del siglo XX se encuentran antecedentes de evaluación en el campo de la educación y la salud pública. El Poder Legislativo ha contado con el Servicio de Investigación del Congreso (Congressional Research Service, CRS), creado en 1914, que analiza las políticas y sus impactos en apoyo a los legisladores. Después de la Segunda Guerra Mundial se realizaron numerosos estudios empíricos sobre programas sociales que utilizaron técnicas de investigación y métodos estadísticos propios de las ciencias sociales.

En la década de 1960, en el marco de la Guerra contra la Pobreza, se implementaron programas que fueron evaluados de manera sistémica tanto para la rendición de cuentas gubernamental como para apoyar la asignación de financiamiento por el Congreso. Hacia el fin de la década se produjeron

* Doctora en ciencias políticas y de la administración, Universidad Complutense de Madrid, estancia posdoctoral en curso, UAM-Xochimilco, México.

¹ Para profundizar, véanse Ballart (1993) y Cardozo (2012a).

estudios experimentales de gran envergadura y elevados costos. Al mismo tiempo se desarrolló el Sistema de Planeación, Programación y Presupuestación (Planning, Programming and Budgeting System, PPBS) que desató procesos de evaluación *a priori*.

En la década de 1970 se crearon y modificaron diversas organizaciones vinculadas con el trabajo evaluativo: la Oficina de Gestión y Presupuesto (Office of Management and Budget, OMB) creada en 1970; la Oficina de Evaluación Tecnológica (Office of Technology Assessment, OTA), creada en 1972 y desaparecida en 1995; la Oficina de Contabilidad General (General Accounting Office, GAO), responsabilizada en 1974 de la evaluación de programas; y la Oficina de Presupuesto del Congreso (Congress Budget Office, CBO), creada en 1975. También se inició la formación universitaria en evaluación y aparecieron las primeras revistas especializadas: *Evaluation News*, actualmente *American Journal of Evaluation*; *Evaluation Quarterly*, actualmente *New Direction for Evaluation and Evaluation and Program Planning*.

En la década de 1980 la GAO creó el Instituto para la Evaluación de Programas (Institute for Program Evaluation, IPE), posteriormente convertido en la División de Metodología y Evaluación de Programas, que facilitó la realización de trabajos de mayor envergadura para el Congreso. Además, se crearon las primeras asociaciones profesionales como el Comité sobre Estándares para la Evaluación de la Educación (Joint Committee on Standards for Educational Evaluation).

En 1993 la Ley de Resultados y Gestión del Gobierno exigió a los organismos públicos que sus objetivos se orientaran a los resultados; mientras en 1998 la GAO se reestructuró para responder a la solicitud del Congreso de realizar auditorías de rendimiento y evaluaciones más complejas. Ese mismo año se aprobó la Ley Federal (Federal Sunset Act) que exige la revisión periódica de la eficacia y eficiencia de las agencias federales, realizadas por una comisión *ad hoc*, que trabaja con la GAO, la OMB y los comités del Congreso.

Institucionalización de la evaluación

Entre los avances a nivel exclusivo del Poder Ejecutivo es de resaltar que, en 2002, la OMB presentó una nueva herramienta para evaluar la eficacia de los

programas federales, llamada Instrumento de Calificación de la Evaluación de Programas (Program Assessment Rating Tool, PART) y ésta empezó a utilizarse por cada agencia y con asesoría de la propia OMB para el Presupuesto Fiscal Federal de 2004. La misma no incluía la evaluación de impacto pero permitía evaluar y observar los factores que afectan el desempeño de los programas e identificar sus fortalezas y debilidades a fin de hacerlos más eficaces y disponer de información para las decisiones de financiamiento y gestión. Se basó en un cuestionario estandarizado que cubría cuatro categorías: el objeto y diseño del programa, la planificación estratégica, la gestión del programa y los resultados; estos últimos eran traducidos en una calificación numérica para mostrar el nivel de cumplimiento de los objetivos anuales y a largo plazo del programa, eran publicados en línea (ExpectMore.gov) y podían llevar incluso a la cancelación del programa. La OMB también trabajaba con cada agencia a fin de preparar planes de mejora para realizar un seguimiento de los hallazgos identificados con el objetivo de mejorar el desempeño de los programas.

Entre 2004-2008 la PART evaluó 98% de los programas del gobierno, que representaba casi 90% del presupuesto federal. Pero, en 2009, el gobierno de Barack Obama cuestionó su uso y la desapareció, principalmente por su falta de interacción con el Congreso. Se anunció entonces una nueva política del gobierno sobre la evaluación y su metodología y un instrumento que reemplazaría a la PART, con énfasis en el uso de la información del desempeño para gestionar programas, la que se publicó en 2010: la Ley de Desempeño Gubernamental y Resultados de Modernización (Government Performance and Results Act Modernization Act,² GPRAMA).

También en 2009 se anunciaron planes para examinar anualmente un conjunto seleccionado de programas en un marco propio de gestión por resultados. En el presupuesto 2011 el gobierno reveló parte de su agenda de diseño, que incluye un requerimiento de metas de desempeño de alta prioridad y énfasis en evaluaciones de programas estrictas e independientes, así como un trabajo conjunto con las agencias para que se disponga de evaluaciones de impacto en línea. La OMB (2010) anunció que otorgaría financiamiento a las agencias federales que estuvieran dispuestas a realizar evaluaciones voluntarias de alta prioridad en sus programas clave.

² Es una reforma de la Government Performance and Results Act (GPRA) de 1993.

En cuanto al Poder Legislativo, como ya se mencionó, tanto el Senado como la Cámara de Representantes del Congreso realizan tareas de evaluación mediante dos instancias técnicas, independientes y no partidistas: la GAO y CBO. La CBO realiza evaluaciones *ex-ante* que incluyen estudios e informes sobre asuntos de políticas con referencia especial a problemas económicos y presupuestarios (sus informes aparecen en línea). La GAO constituye el brazo auditor, evaluador y de investigación del Congreso; se encarga de examinar el uso de los fondos públicos, realizar evaluaciones de programas y análisis de políticas públicas, efectuar recomendaciones y apoyar al Congreso para elaborar políticas y fundamentar decisiones. Tiene su oficina central en Washington, DC., cuenta con 11 oficinas regionales y se financia con asignaciones presupuestales del Congreso.

En julio de 2004 la GAO cambió su nombre a Oficina de Contabilidad Gubernamental (Government Accountability Office) con el fin de clarificar la verdadera función que está cumpliendo: en este año las auditorías financieras presentaban sólo 15% de su trabajo, la mayoría del mismo involucraba evaluaciones de programas, análisis de políticas, así como opiniones y decisiones legales en un amplio rango de programas y actividades del gobierno, tanto a nivel nacional como en el extranjero.

La GAO logra mantener su independencia debido a que el contralor general (nombrado por el presidente de los EUA a partir de una lista de candidatos propuesta por el Congreso) desempeña su cargo durante 15 años, pudiendo ser removido sólo por el voto del Congreso. La independencia también es reflejo de su personal, integrado por profesionales del mundo académico, del sector privado, y de otras agencias gubernamentales. Unas tres cuartas partes de sus empleados trabajan como evaluadores, analistas y especialistas técnicos o de campos específicos. También tiene auditores especializados (70% de ellos con nivel de doctorado o maestría) en materias como el medio ambiente, arquitectura, ingeniería, estadística, sanidad, nuevas tecnologías, etcétera. Si bien las evaluaciones son realizadas por el personal de la GAO, ésta está autorizada a contratar expertos externos para consulta y asesoría.

La GAO puede actuar por mandato legislativo y peticiones específicas de comisiones del Congreso, por requerimiento de un congresista, por iniciativa propia, o porque así lo establece la normatividad. Esta última está referida a la planificación, al alcance de una auditoría, al diseño de métodos de recopilación

y análisis de datos, a las formas de documentar las decisiones y los modos de informar sobre los resultados de trabajo, etcétera, en el llamado “Libro Amarillo”.

El espectro de trabajo de la GAO incluye actualmente todo lo que el gobierno federal está haciendo o pensando hacer en cualquier lugar del mundo (i.e. conflicto en Irak). Entre los reportes más recientes que aparecen en su página web se incluyen temas como los mercados de la electricidad, regulaciones ambientales, prevención de la salud, desempeño de los contratistas, programas de alimentos escolares, transparencia de datos, viviendas prefabricadas, fondos para el trabajo de la fuerza aérea, sólo por mencionar algunos. Como puede verse, aunque gran parte de su trabajo está dedicado a revisar la eficacia de las actividades gubernamentales, también atiende el curso de los acontecimientos internacionales y problemas emergentes, en estrecha colaboración con el Congreso.

Hoy, gran parte de las evaluaciones más importantes de la GAO se traducen en informes guiados por valores de responsabilidad, integridad y fiabilidad sobre programas, asuntos de “alto riesgo” y “mejores prácticas”. Como “alto riesgo” identifica y analiza los programas nacionales clave que conllevan un riesgo inherente en términos de fraude, generación de residuos, abusos, ineficiencia en los sistemas de gestión, posibilidad de fracaso de algún programa importante, pérdida de vidas humanas, afectación de la salud o la seguridad. En 2013 se ubicaron 30 áreas problemáticas de alto riesgo.

Las “mejores prácticas” se refieren al análisis de las organizaciones públicas y privadas ampliamente reconocidas por las mejoras experimentadas en el rendimiento en algún área específica. Sus procesos, prácticas y sistemas sirven de modelo a otras organizaciones que tienen una función o misión similar, como parte de la preocupación de la GAO de cooperar para la actualización y modernización de la administración pública.

Además de diseñar y realizar evaluaciones de programas, el personal de la GAO también ha perfeccionado y difundido metodologías de evaluación. En este sentido, ha publicado numerosos manuales y guías de evaluación que son aplicados por muchos organismos públicos y privados, además de varios informes sobre estrategias, enfoques y cultura de la evaluación de programas. Los resultados de su trabajo (evaluaciones e informes) se publican tanto en su página institucional como a través de las asociaciones profesionales de evaluación. La GAO también mantiene en su página web un blog sobre eventos de actualidad y realiza un *podcast* con entrevistas a sus oficiales.

Para concluir este apartado, resulta necesario referirse también al trabajo de las consultorías,³ tanto en la realización de evaluaciones como en la mejora de los programas de las organizaciones en función de sus resultados. Suelen ofrecer sus servicios tanto a organizaciones públicas como privadas, así como a organizaciones sin ánimo de lucro y fundaciones. Algunas no sólo trabajan dentro de los EUA, sino también a nivel internacional. En el ámbito público, los clientes son generalmente los departamentos y agencias de gobiernos locales. Las áreas son diversas y se incluye, por ejemplo, el transporte, el comercio o la energía, pero las más numerosas pertenecen a las áreas sociales, de educación y de salud. Sus servicios se basan en su mayoría en enfoques participativos y colaborativos, por lo que desarrollan sus actividades en estrecha colaboración con sus clientes. En algunos casos, sus miembros han publicado libros sobre el tema.

Trosa (2003) recupera muchos de los aspectos antes mencionados al caracterizar al sistema de EUA de la siguiente manera: existe una cultura nacional de evaluación y ésta es utilizada en el debate político tanto por el Estado Federal como por los Estados federalizados, el Parlamento o los *lobbies* para defender sus proyectos o para argumentar en contra de otros; se trata del país que dedica más recursos a la evaluación (mil personas en la GAO); cuenta con múltiples instituciones públicas y privadas dedicadas a la recolección de información y publicación de estadísticas; la promoción de la evaluación es encabezada por la OMB y el Parlamento, lo que refuerza el papel de este último; su objetivo es discutir la pertinencia de los programas y cuenta con el más alto grado de formalización; su nivel de independencia es importante aunque 80% son combinaciones de trabajos internos/externos; y el control de calidad sólo está asegurado en los casos en que son examinados por la Oficina de Contabilidad Gubernamental.

³ Es el caso de EvalSolutions, Gargani + Company, Fulcrum Corporation, JBS International Burlingame, Fatterman & Associates: Evaluation Consultation Services, Community Evaluation Solution y The Evaluation Center of the Western Michigan University, cuyas páginas fueron revisadas.

Profesionalización de la evaluación

Asociaciones profesionales

Cuando la demanda de evaluaciones aumentó, sus profesionales se agruparon para conformar la Sociedad de Investigación en Evaluación (Evaluation Research Society) y la Red de Evaluación (Evaluation Network), que posteriormente se unirían para formar la actual Sociedad Americana de Evaluación (American Evaluation Association, AEA). Estas asociaciones han influido fuertemente en el desarrollo de la evaluación.

La AEA integra académicos y consultores, tanto de los EUA como de otros países. Está comprometida con la aplicación e investigación de la evaluación de programas, de personal, de tecnología, entre otras. Su misión es mejorar los métodos y prácticas en la materia, incrementar su uso, promoverla como una profesión y apoyar su contribución a la generación de teoría y conocimiento sobre la acción humana. Contribuye a la enseñanza y aprendizaje de la evaluación entre sus miembros; en el desarrollo y difusión de nuevo conocimiento; en la promoción de la comunicación y el uso significativo de la evaluación por los *stakeholders*; y en la interacción multicultural con diversas comunidades para lograr trabajos eficaces y respetuosos, al mismo tiempo que fomenta la integración cultural, geográfica y económica, la justicia social y la igualdad. Promueve estándares y principios de evaluación; mantiene vínculos de comunicación a través de su página web; entrega premios que reconocen la excelencia en el campo de la evaluación; ofrece diversos programas y servicios a estudiantes y profesionales; además de que publica las revistas *American Journal of Evaluation* y *New Directions for Evaluation*.

Esta Asociación permite al público en general acceder a sus recursos, pues intenta que éstos sirvan para conocer las diversas formas de hacer y usar la evaluación; mantener puntos de vista positivos y realistas sobre los roles que ésta puede tener; hacer de ella una práctica común en la operación de sus organizaciones; y ser sensibles a las cuestiones éticas que involucra el realizarla y utilizarla. También busca influir en la configuración y uso de la política de evaluación de los Estados Unidos de América.

Además ha elaborado lineamientos que guían la práctica evaluativa; la más relevante es la “Guía de principios para los evaluadores”, publicada inicialmente

en 1994 y revisada en 2004. La “Declaración de la competencia cultural en evaluación”, de 2011, es otra guía dirigida a todas las evaluaciones que se pretenden realizar alrededor del mundo. En 2013 publicó también una “Hoja de ruta para un gobierno más efectivo”, que describe su visión sobre el rol de la evaluación en el gobierno federal y detalla un plan para mejorar su calidad.

Dentro de la Asociación se han conformado 51 grupos de interés de actualidad en los que se discuten temas relevantes relacionados con la evaluación, como puede ser su uso, la pluralidad, la metodología, los métodos o el aprendizaje organizacional. Estos grupos discuten temas de políticas sectoriales variadas que van desde la salud o la educación hasta la economía, pasando por asuntos feministas, poblaciones vulnerables, juventud o gestión de desastres.

Otro espacio para mantener contacto con evaluadores y temas de interés es la sección “AEA365. Un consejo al día por y para evaluadores”. Surgió en enero de 2010 y su propósito es que los evaluadores de todo el mundo contribuyan diariamente con un comentario relativo al trabajo que han venido desarrollando y que sea de interés para el resto de la comunidad. La contribución no debe ser extensa ni muy formal, comienza con una breve presentación personal seguida del relato de su trabajo.

Una de las actividades más relevantes de la AEA son sus conferencias anuales que se realizan desde 1986. Éstas reúnen a evaluadores de EUA y del resto del mundo para compartir sus experiencias profesionales. Cada año se selecciona un tema de actualidad a destacar durante el evento, por ejemplo, la utilización de la evaluación, su componente político, el aprendizaje, el fortalecimiento de su capacidad, la incorporación de la perspectiva de evaluación (*Mainstreaming evaluation*), el contexto, la calidad de la evaluación o los valores. Los temas en 2013 y 2014 se concentraron bajo los títulos de “El estado de la evaluación en el siglo XXI” y “Evaluación visionaria para un futuro sostenible y equitativo”, respectivamente. En 2014 el propósito ha sido vincular a la evaluación con el principal desafío actual: el desarrollo de un mundo diverso e interconectado que resulte sustentable para la generación actual y las venideras. Ya que en la intersección de los asuntos medioambientales, de equidad, y económicos, existen oportunidades para revisualizar la evaluación y su contribución a un mundo sostenible y equitativo, aunado a la importancia que los temas de transparencia, rendición de cuentas y eficacia han alcanzado alrededor del mundo.

Publicaciones especializadas

Al igual que las asociaciones profesionales, las publicaciones especializadas también constituyen esfuerzos colectivos dirigidos a mejorar la evaluación y han contribuido a delinear su campo. En materia de libros, la publicación es vasta; a continuación se mencionan los más relevantes publicados de 2008 a la fecha, clasificados en cuatro grupos.

- a) Libros que abordan temas básicos sobre evaluación y/o guías para una buena práctica. En 2008 se publicaron dos. El primero abarca los aspectos esenciales en evaluación de programas (Spaulding, 2008), incluyendo sus tipos, fundamentos, herramientas para la recolección de datos, redacción de reportes, y comunicación de hallazgos, mientras el segundo, identifica las competencias y estrategias necesarias para quienes conducen evaluaciones en un marco organizacional, principalmente dinámico y cambiante (Russ-Eft *et al.*, 2008). Ambos libros incluyen casos prácticos. Al siguiente año se publicaron dos manuales, el primero es una guía para la planificación completa y factible de la evaluación en cualquier contexto, que ofrece herramientas conceptuales para que ésta sea exitosa, útil e informativa (Holden y Zimmerman, 2009). El segundo presenta los métodos de evaluación y sus aplicaciones, particularmente en la formulación de políticas y toma de decisiones (Ryan y Cousins, 2009). Posteriormente se publicó un tercer manual, enfocado en el desarrollo de una evaluación en las áreas de trabajo social y salud, tomando en cuenta la importancia de perspectivas actuales como el empoderamiento del consumidor y la evidencia basada en la práctica (Smith, 2010).
- b) El segundo grupo incluye las publicaciones relativas al uso de las evaluaciones, donde Michael Patton refleja su autoridad en el tema. Uno de sus libros enfatiza el trabajo colaborativo con los tomadores de decisiones de los programas, especialmente en el caso de evaluaciones del desarrollo, basándose en ideas sobre sistemas dinámicos, complejos y emergentes (Patton, 2011). Publicado por primera vez en 1978, la nueva edición de *Essentials of Utilization-focused Evaluation* (Patton, 2012) incorpora los más recientes cambios en el campo de la evaluación, presentando los avances en la teoría y práctica, basados en la investigación emergente sobre la evaluación y su uso.

- c) En el tercer grupo se ubican los textos referidos a la metaevaluación. Aunque no es extensa la producción de material sobre el tema, sus referencias se ubican dentro de los principales glosarios y enciclopedias que exponen sus propósitos y beneficios, la forma de desarrollarla y sus consideraciones teóricas y metodológicas (Schroeter, 2009 y Wingate, 2010). A título de ejemplo, se identificó una tesis doctoral no publicada (Wingate, 2009).
- d) Finalmente, los libros sobre una gran variedad de temas de interés actual en evaluación, cuyo número apenas permite mencionarlos. Una publicación pone en perspectiva el significado de la evidencia (Donaldson, Christie y Mark, 2009); otra aborda la complejidad creciente de los contextos de evaluación (Morell, 2010); una más analiza la teoría que sustenta el programa (Funnell y Roger, 2011). En 2012 los trabajos se enfocaron en el aspecto colaborativo de la evaluación (Cousins y Chouinard, 2012) y se publicó una guía para evaluadores que buscan tener éxito en la consultoría independiente (Barrington, 2012). Otro libro abordó el aspecto interpersonal de la práctica evaluativa (King y Stevahn, 2013) y uno más ofreció una guía del proceso de monitoreo y evaluación en un marco democrático (Kumar, 2013). Por último, se realizó una nueva edición de un libro sobre la evaluación para el empoderamiento (Fetterman, Kaftarian y Wandersman, 2014).

Cuando las revistas especializadas comenzaron a tomar relevancia decidieron adoptar la revisión por pares y se convirtieron en revistas más académicas. A continuación mencionamos las principales temáticas y trabajos publicados en los últimos años en las revistas más destacadas.

La *American Journal of Evaluation* analiza las decisiones y los desafíos relacionados con la conceptualización, el diseño y la realización de evaluaciones. Publica artículos sobre los métodos, las teorías, la ética, la política y la práctica de la evaluación, con una perspectiva multidisciplinaria y referidos a áreas como educación, administración pública, ciencias del comportamiento, servicios humanos, ciencia de la salud, sociología, criminología, etcétera. También ha dedicado espacio a temas como la evaluación participativa, las tensiones entre evaluación e investigación, la presentación de los resultados, la ética, el rol de los *stakeholders* o la evaluación formativa. En 2001, 2005 y 2008 publicó artículos sobre metaevaluación. Recientemente ha organizado los foros de discusión denominados Creación de Capacidad en Evaluación (Evaluation Capacity

Building, ECB), sobre las perspectivas internacionales, la metodología y la evaluación en el ciclo de vida de los programas.

La *Evaluation Review* ha servido como foro interdisciplinario para los investigadores de las ciencias sociales, los planificadores, y los tomadores de decisiones de políticas dirigidas a mejorar la condición humana. Los artículos publicados giran principalmente en torno a los métodos de evaluación usados en disciplinas como educación, salud pública, justicia penal, desarrollo infantil, salud mental, trabajo social, administración pública y medio ambiente.

New Directions for Evaluation incluye trabajos empíricos, metodológicos y teóricos sobre evaluación. Cada número está dedicado a un tema en específico, por ejemplo, actuación del gobierno, energía, medio ambiente, salud mental, educación, desarrollo de recursos humanos, evaluación de resultados, evaluación del personal, análisis de políticas o evaluación de tecnología. Uno de sus números de 2012 abordó el tema de la evaluación en América Latina; mientras en 2014 otro número se dedicó a la capacidad organizacional para realizar y usar la evaluación.

Evaluation and Program Planning se basa en el principio de que las técnicas y métodos de evaluación y planeación trascienden las barreras de los campos específicos, las tradiciones intelectuales y los intereses. La revista publica artículos tanto del sector público como del privado en diversas áreas: desarrollo y comportamiento organizacional, capacitación, planeación, desarrollo de recursos humanos, salud, servicios sociales y educación. Presenta tanto reportes de evaluaciones específicas y esfuerzos de planeación, como debates de temas relevantes. En los últimos años, entre otros, se han publicado artículos sobre la utilización de la evaluación y sobre procesos participativos y evaluación colaborativa.

La revista de más reciente creación es *Journal of Multidisciplinary Evaluation*, que surge en 2004. Michael Scriven fue el encargado de introducir el primer número en el que sostenía la importancia de la tarea de comunicar los avances en evaluación así como sus usos. Los primeros artículos de la revista se enfocaron a los procesos participativos, pero otro de sus principales propósitos fue el de realizar revisiones periódicas sobre lo que sucedía en materia de evaluación fuera de los EUA, por lo que dedicó varios números a una revisión global de la situación en diversas partes del mundo: África, Europa Occidental, Europa del Este y Medio Este, América Latina y el Caribe, Asia y Oceanía. En 2009 dedicó cuatro

artículos a la teoría, los métodos y la práctica de la metaevaluación. Su editorial estuvo a cargo de Scriven (2009) y abordó la definición, justificación y aplicación de esta actividad. En 2011 se publicaron artículos de autores reconocidos dentro del ámbito (Worthen y Sanders, 2011 y Sanders y Nafziger, 1976), pero en los últimos dos años la revista ha reducido el número de sus artículos.

En torno de la metaevaluación, del 2000 a la fecha se ubicaron sólo alrededor de diez artículos, lo que refleja su lento desarrollo. Éstos abarcan tanto discusiones teóricas como estudios de casos. Destacan los que publica la revista *Journal of Multidisciplinary Evaluation*; donde Cooksy y Caracelli (2009) analizan 18 evaluaciones que informan a los *stakeholders* acerca de sus fortalezas y debilidades, mientras LaBelle (2009) intenta determinar el impacto efectivo de las evaluaciones utilizadas para la mejora de los programas en su artículo, y Hedler y Gibram (2009) se refieren a la diferencia entre metaanálisis y metaevaluación, realizan un modelo para la última y presentan los resultados de un estudio empírico realizado en la agencia de auditoría del gobierno de Brasil. Esta revista también rescata el artículo “Metaevaluation” (Stufflebeam, 1974) que aborda su metodología, analizando los factores y problemas asociados con su realización, la necesidad de llevarla a cabo, los criterios que deberían guiarla, los problemas que ponen en riesgo la evaluación y deben ser abordados en la metaevaluación, además de presentar una estructura lógica para diseñar un estudio de este tipo. Un artículo de Nkwake, Trandafilí y Hughey (2013) se acerca más bien a efectuar un metaanálisis de las evaluaciones realizadas por una organización no gubernamental entre 2005 y 2010.

Por su parte, la *American Journal of Evaluation* publicó un artículo más de Stufflebeam (2001), en el que resalta el avance de la evaluación y la necesidad de que ésta se someta a metaevaluación. En 2005 aparece un artículo que aborda las múltiples formas en que se puede definir la calidad, los contextos políticos y sociales de la metaevaluación y las cuestiones relacionadas con su uso (Cooksy y Caracelli, 2005). Además, sus autoras exponen una metaevaluación de las evaluaciones de los centros internacionales de investigación agrícola pública. Se publica también un artículo en el que Hanssen *et al.* (2008) abordan la realización de metaevaluaciones simultáneas con la evaluación, ya que ofrecen la oportunidad para que los metaevaluadores asesoren a los evaluadores.

Respecto del uso de los resultados y las recomendaciones de la evaluación, la producción ha sido mucho más extensa, lo que refleja su importancia. Debido

a su número sólo se mencionan los temas variados que se abordan: uso de la evaluación en el contexto organizacional, marcos para la negociación, enfoques para la presentación de resultados, modelos conceptuales para comprender el impacto de la evaluación en el sector público, enfoques que contribuyen a su utilización o importancia de la metodología para promover su uso, sólo por mencionar algunos.

Programas académicos

Éstos pueden abarcar posgrados, cursos, seminarios, prácticas o diplomados diseñados para enseñar los principios de la evaluación y formar profesionales en el área. LaVelle y Donaldson (2010) realizaron en 2008 un estudio para conocer la oferta de estos programas en los EUA y encontraron un total de 48, la mayoría de ellos localizados en escuelas de educación (60%), psicología de la educación (14.5%), psicología (10.5%), política pública (8%), justicia criminal (2%), sociología aplicada (2%) y estudios interdisciplinarios (2%). De éstos, 18% ofrecían sólo el nivel de maestría, 29% sólo el de doctorado y 52% ambos niveles. Los resultados de su estudio muestran un panorama consistente de la disciplina y profesión de la evaluación de cara al crecimiento de esta demanda. Aunque también sugieren que la evaluación todavía está fuertemente unida a sus disciplinas hermanas y a menudo integrada con métodos de investigación y análisis de datos, además de que existen tipos poco estudiados, como la metaevaluación.

La academia realiza contribuciones importantes a la evaluación, como ayudar a entender los vínculos entre la teoría y la práctica, conocer lo que se considera una evidencia creíble, mejorar las relaciones entre evaluadores y *stakeholders*, desarrollar las competencias esenciales para la práctica y establecer estándares o guías de principios para los evaluadores (LaVelle y Donaldson, 2010). Estos avances son muy importantes en la medida en que la disciplina busca posicionarse en el panorama profesional, lo que implica ofrecer una formación que domine las habilidades específicas de la evaluación para brindar servicios de calidad a los clientes.

En Washington se encuentra The Evaluator's Institute que forma parte de la Universidad George Washington y constituye la primera entidad del

mundo dedicada exclusivamente a formar profesionales en evaluación. Ofrece una gran variedad de cursos en el tema, divididos en cinco categorías: *a*) enfoques analíticos; *b*) enfoques y técnicas de evaluación; *c*) fundamentos de la evaluación; *d*) evaluación, teoría, diseño y métodos; y *e*) uso de la evaluación (estrategias y capacidad). Cada categoría contiene al menos nueve opciones de cursos y dependiendo del número de cursos tomados se obtiene un cierto nivel de certificación, donde el cuarto es una maestría.

En el Cuadro 1 se presentan los programas de doctorado y maestría que ofrecen las universidades de los EUA relativos a la formación profesional de evaluadores, así como cursos adicionales a los ya mencionados en The Evaluator's Institute. Debido a su amplia oferta sólo se incluyeron los referidos a la evaluación de políticas y programas públicos en general, dejando fuera los más amplios, como los programas en políticas públicas, o los más específicos, como los dedicados a la evaluación en áreas concretas como la salud, la psicología, la educación o el trabajo social. La mayoría de los programas se encuentran en escuelas de política pública o de administración.

La práctica de la evaluación cuenta con una larga tradición en Estados Unidos, al día de hoy está plenamente institucionalizada, por lo que es un referente mundial. Dentro del gobierno, tanto el Ejecutivo como el Legislativo poseen atribuciones en la materia y la consultoría también tiene una presencia importante. Se ha creado una fuerte asociación de profesionales de la evaluación, se publican numerosos libros y varias revistas especializadas en el tema. La metaevaluación, aunque poco recurrente, también está presente. Además, aunque todavía aparece muy ligada a métodos de investigación y análisis de datos, EUA cuenta con una amplia oferta de programas universitarios, e incluso con institutos especializados, que forman evaluadores de cara al crecimiento de la demanda y la calidad de la práctica.

CUADRO 1

Estados Unidos: principales programas académicos por universidades

Programa	Universidad/Institución	Descripción
Doctorado en evaluación	Universidad Western Michigan	Desde 2004 ofrece un doctorado interdisciplinario para abordar la creciente necesidad y demanda de alto entrenamiento y habilidad en los evaluadores.
Doctorado en evaluación de programas	Universidad George Washington, Escuela de Política Pública y Administración	Desarrolla habilidades interdisciplinarias de investigación cuantitativa y cualitativa aplicadas para evaluar programas públicos y no gubernamentales.
Doctorado en evaluación y medición	Universidad del Estado de Kent, Facultad de Educación, Salud, y Servicios Humanos	Ofrece un curso sobre evaluación de programas así como de diseños de investigación tanto cuantitativo como cualitativo.
Doctorado en diseño instruccional, desarrollo y evaluación	Universidad de Syracuse, Escuela de Educación	Tiene un módulo exclusivo de evaluación y métodos de investigación formado por seis áreas.
Doctorado en evaluación y métodos de investigación aplicada; maestrías en evaluación de programas: política pública, educación y psicología social	Universidad Claremont, División de Ciencias del Comportamiento y Organización	Desde la década de 1960 ha formado profesionales, muchos de los cuales se desempeñan en el ámbito de la evaluación y la investigación aplicada en el gobierno.
Doctorado y maestría en metodología de la investigación social	Universidad de Los Ángeles, California, Departamento de Educación	Enfocado en la investigación educativa y social, así como en la investigación y evaluación.
Doctorado y maestría en medición, estadística y evaluación	Universidad de Maryland	Dirigido a los interesados en los métodos de investigación y su aplicación a la evaluación.
Maestría en evaluación de programas	Universidad Duquesne, Escuela de Educación	Diseñada para incrementar el conocimiento teórico de la evaluación y mejorar las habilidades en el manejo de sus métodos.
Maestría en evaluación de programas	Universidad del Estado de Michigan	Ofrece conocimiento teórico y experiencias reales para que el estudiante trabaje en un entorno de evaluación junto a un profesional de la misma.
Cursos sobre auditoría y evaluación de programas y políticas públicas	Escuela de Auditoría de la Oficina de Contabilidad Gubernamental	Se enfoca en la enseñanza y perfeccionamiento del personal de la propia Oficina de Contabilidad Gubernamental a partir de clases prácticas y teóricas relacionadas con su trabajo.

Fuente: páginas web de las respectivas universidades, revisadas en agosto de 2014.

CAPÍTULO II

Reino Unido

Talina Olvera Mejía

El Reino Unido ha sido un referente en la aplicación de las reformas propuestas por la Nueva Gerencia Pública (*New Public Management*, NPM) y en el uso de herramientas de gestión por resultados para mejorar la eficiencia, lo que necesariamente lo ha llevado a incursionar en la evaluación de políticas y programas públicos. Antes de abordar el estado actual de este tema, se presenta una breve síntesis de lo realizado por este país durante la segunda mitad del siglo XX.¹ A finales de la década de 1960 y principios de la de 1970 el interés por el análisis de las políticas y la evaluación retrospectiva de sus resultados se apoyó tanto en las iniciativas del Tesoro (Ministerio de Economía y Finanzas del gobierno), para poner en funcionamiento sistemas efectivos de control de gasto, como en el movimiento de reforma que reclamaba mejor gestión en el sector público y más racionalidad en los procesos de decisión. No obstante, las iniciativas creadas para estos propósitos desaparecieron sin haberlos conseguido.

Con el apoyo de Margaret Thatcher en el poder, en 1979 la principal preocupación fue el control de los recursos utilizados. Se diseñó la Estrategia de Eficiencia (Efficiency Strategy) para mejorar su utilización y la Iniciativa de Gestión Financiera (Financial Management Initiative) para ampliar la revisión de políticas por parte del Parlamento y la opinión pública. Pero nuevamente, la excesiva preocupación por los recursos y el establecimiento de

¹ En este apartado se retomaron los aportes de Ballart (1993) y Cardozo (2012a).

sistemas de control de gestión internos dejó poco margen a la evaluación de la eficacia de las políticas.

A nivel parlamentario, en 1983 la Comisión de Contabilidad Pública reforzó su capacidad de control mediante la aprobación de una ley que amplió las funciones tradicionales de la Oficina Nacional de Auditoría (National Audit Office, NAO), incluyendo facultades para investigar si los departamentos del Poder Ejecutivo utilizaban los recursos de forma eficiente y eficaz. Aunque difícilmente estas investigaciones fueron evaluaciones en sentido estricto.

Después de un intenso proceso de capacitación en el gobierno, en 1988 se publicó el manual *Policy Evaluation. A guide for managers* para apoyar a los altos funcionarios en el desarrollo de evaluaciones de políticas públicas. Su principal propósito fue promover la evaluación, desmitificando su contenido y tratando de reducir los temores que generaba.

Las reformas de la década de 1980 y la mayor parte de la de 1990 se centraron en la optimización y la eficiencia por medio del concepto *value for money*.² En 1991, existía un avance muy desigual de la integración de la evaluación en los diferentes sectores de la actividad gubernamental, por lo que al año siguiente el Tesoro aumentó su vinculación con el sistema de asignación de recursos. Desde 1994 la Comisión de Auditoría (Audit Commission) exige a las instancias gubernamentales la publicación anual de indicadores no financieros.

El gobierno laborista iniciado en 1997 introdujo un sistema más explícito de planificación y presupuesto por resultados, *Best Value*, que combina el enfoque racional-objetivo de la evaluación de resultados e impactos con el argumentativo-subjetivo dirigido al aprendizaje organizacional. Bajo el enfoque *evidence-based policy* la formulación de políticas y las decisiones se deben fundamentar en el conocimiento de lo que funciona bien (*what matters is what works*), siendo indispensable apoyarse en evidencia científica o de investigación y en evaluaciones rigurosas de políticas.

² *Value for money* se ha definido como la demostración de una adecuada gestión y generación de resultados al menor coste posible, alcanzando un elevado nivel de satisfacción de los usuarios. No constituye una simple medida de los costes, sino que engloba una combinación de variables como calidad, coste, uso de recursos, idoneidad, oportunidad y conveniencia, para juzgar si, en conjunto, se aporta un nivel de valor adecuado en la prestación del servicio (Pontones y Pérez, 2013).

Institucionalización de la evaluación

Actualmente, además del énfasis en la gestión por resultados, se busca poner más atención a la consistencia del conjunto de políticas gubernamentales. Pero, pese al nivel de institucionalización alcanzado por la evaluación desde 1997, Talbot (2010) considera que el gobierno se ha orientado más hacia el enfoque del monitoreo del desempeño organizacional (*performance*) que a la evaluación de políticas y programas.³

Dentro de las ya mencionadas reformas del gobierno laborista se avanzó en la asignación del presupuesto por resultados mediante acuerdos entre los ministerios y el Tesoro. Para incrementar la flexibilidad en la utilización del presupuesto los ministerios firman los Acuerdos de Servicio Público (*Public Service Agreements*) en que se comprometen a obtener los objetivos y metas acordados para mejorar los servicios públicos y sus efectos (*outcomes*). Sus logros y costos son revisados posteriormente en función de las prioridades del gobierno a partir de las Revisiones de Gastos (*Spending Reviews*), que constituyen una especie de evaluación del desempeño.

Las evaluaciones y revisiones pueden ser muy variadas en su objetivo y estar coordinadas por las agencias centrales de presupuesto como el Tesoro, por los ministerios sectoriales o por la NAO. La Oficina del Gabinete y el Tesoro sugieren metodologías diversas para realizarlas y los departamentos eligen las que consideran pertinentes. Éstos las aplican, en parte porque el sistema de gestión por resultados así lo exige, pero también porque existe una creciente cultura gerencial. En el sitio web del gobierno se puede acceder a los documentos por departamentos y agencias, incluyendo los informes de evaluación. La mayoría de ellos han desarrollado sus propios procesos de control de la calidad de las

³ Además de que el monitoreo se centre en las organizaciones y la evaluación en las políticas o programas, también influye el tiempo, los métodos y los solicitantes. El primero se realiza de manera periódica, generalmente anual; utiliza principalmente el método cuantitativo y es solicitado por los encargados de finanzas, operaciones o incluso por las ramas de personal. En cuanto a la evaluación, éstas son episódicas, usualmente únicas o excepcionales; su metodología es mixta –tanto cualitativa como cuantitativa–, y son solicitadas en su mayoría por los encargados de las ramas de políticas.

evaluaciones, ya que el gobierno central no cuenta con mecanismos para su revisión sistemática.

El Tesoro, junto con otras áreas del gobierno, también puede encargar revisiones *ad hoc* de políticas a expertos, principalmente cuando el tema es altamente técnico o polémico. Una de las agencias con las que trabaja es la Oficina de Responsabilidad Presupuestaria (Office for Budget Responsibility),⁴ encargada entre otras funciones de juzgar la actuación del gobierno sobre sus objetivos fiscales.

Por su parte, cada departamento gubernamental es responsable de realizar evaluaciones *ex-ante* y *ex-post* con base en metodologías centralmente establecidas y de reportar sus resultados a la oficina del primer ministro. Pueden realizarse internamente o contratar unidades de evaluación independientes. Los departamentos tienen la responsabilidad de producir un informe de evaluación conforme con una guía de procedimientos y presentación de resultados en que se revisan las evidencias basadas en fuentes específicas, se analizan los resultados de políticas experimentales y se complementan con los resultados de otras evaluaciones. También realizan evaluaciones económicas, estudios comparativos, auditorías y evaluación del desempeño de gestores.

Para lograr que el uso de recursos produzca el mayor bienestar social y económico el gobierno requiere de políticas basadas en evidencia fiable, por lo que considera vital la calidad de las evaluaciones. En atención a esto, ha buscado fortalecer la función de evaluación y análisis de políticas públicas a partir de la elaboración de guías de evaluación por parte del Tesoro, como el Libro Verde⁵ y el Libro Magenta⁶ que apoyan a los tomadores de decisiones y analistas en la revisión y valoración adecuada de políticas y proyectos. Estas dos guías son complementarias: el Libro Verde pone énfasis en los principios económicos que deben ser aplicados en la estimación de los costos y el Libro Magenta en el diseño y desarrollo de evaluaciones.

El Libro Verde ofrece orientación sobre la evaluación económica *a priori* del costo y los beneficios de las opciones de políticas. Es un requerimiento para

⁴ Esta Oficina es la equivalente al Congressional Budget Office de los Estados Unidos.

⁵ El Libro Verde fue preparado en 2003 y tuvo una pequeña modificación en 2011.

⁶ El Libro Magenta tiene su versión más reciente en 2011.

todos los gastos y para todas las acciones nuevas de políticas que pueden tener un impacto en los negocios, el voluntariado y el resto del sector público. Describe cómo deberían combinarse las valoraciones económicas, financieras, sociales y medioambientales y contiene guías suplementarias por área.⁷

El Libro Magenta fue desarrollado en el contexto de la demanda de formulación de políticas basadas en evidencias, por lo que inicia señalando el riesgo de no evaluar o de hacerlo mal para los tomadores de decisiones. Establece las mejores prácticas para que sean tomadas en cuenta por los departamentos, gobiernos locales y el sector del voluntariado, buscando promover el entendimiento de los conceptos necesarios y orientando sobre el diseño y la realización de evaluaciones de proceso (valorar si una política se realiza como se había pretendido), de impacto (analizar los cambios que han ocurrido y su posible atribución a la política aplicada) y económicas (comparar los beneficios de una política con sus costos económicos).

Por su parte, el Departamento del Trabajo y Pensiones en conjunto con el Centro Nacional para la Investigación Social han desarrollado el documento “Métodos de investigación para la evaluación de políticas”, usado especialmente en la evaluación de políticas y programas del sector laboral (Talbot, 2010:14).

A nivel parlamentario, de manera totalmente independiente del gobierno, la NAO examina el gasto público en nombre del Parlamento para salvaguardar los intereses de los contribuyentes y ayudar al gobierno a mejorar los servicios públicos mediante una toma de decisiones informada. Selecciona las áreas a auditar mediante un enfoque de gestión de riesgos, tomando en consideración las opiniones del Comité de Cuenta Pública (Public Account Committee), del Parlamento, los Departamentos y otros interesados. Aplica la metodología de *value for money* para auditar los estados financieros de los organismos públicos y realiza evaluaciones *ex-post* de programas y procesos seleccionados dentro de una amplia gama de áreas y temas, pero no tiene la obligación de evaluar las políticas en sí mismas. Los reportes de sus trabajos aparecen publicados en su

⁷ Algunos ejemplos de las guías suplementarias actualizadas en 2013 son las relativas a calidad del aire, crimen, medio ambiente, transporte y salud.

página de internet;⁸ también ofrece guías para realizar auditorías de gestión. La NAO, a su vez, es supervisada por la Comisión Parlamentaria de Contabilidad Pública, que también designa a los auditores externos y controla su desempeño.

Trosa (2003) resume de la siguiente manera las características del sistema de evaluación del Reino Unido que se han venido planteando: existe una cultura nacional de la evaluación ligada más a la eficiencia que a los impactos de la política pública; ha sido impulsada por el Tesoro; persigue el mejoramiento continuo de las políticas públicas y la responsabilización de los funcionarios; está plenamente integrada a la administración desconcentrada de las agencias; constituye el principal pilar de la política de modernización de los servicios públicos tanto a nivel del primer ministro, como del Tesoro o el Parlamento; es independiente en la fase de preparación pero más controlada en el resto del proceso; y la calidad de los trabajos no está oficialmente supervisada, de manera informal puede hacerla la oficina del primer ministro o el Tesoro, pero no se hace pública.

Profesionalización de la evaluación

Asociaciones profesionales

La principal asociación es la United Kingdom Evaluation Society (UKES), fundada en 1994. Busca promover y mejorar la teoría, la práctica, el entendimiento y la utilización de la evaluación, además de fomentar el diálogo y el debate intersectorial e interdisciplinario. Dicha sociedad pone especial atención al crecimiento y desarrollo de las redes de evaluación a nivel regional y nacional en

⁸ Los reportes de los últimos meses, consultados el 15 de septiembre de 2014, tienen que ver con: “Financiamiento al cuidado de la salud”, “Actualización sobre el Programa Nacional de Seguridad Cibernética”, “Transformación de la gestión de contratos de gobierno”, “Participación de 16 a 18 años de edad en la educación y la formación”, “Adquisición de nuevos trenes”, “Programa de trabajo”, “Primeros contratos para la electricidad renovable”, “Esquema 2012 de mantenimiento infantil: primeros avances”, “Mantenimiento de la infraestructura estratégica: caminos”, “Actualización sobre los preparativos para la medición inteligente” y “Servicios de balance de electricidad”.

el país, como una forma de intercambiar información acerca de las oportunidades de evaluación en la región. En este sentido, existen cinco redes regionales: la Red de Evaluación de Londres, la Red de Evaluación del Noreste, la Red de Evaluación del Noroeste; la Red de Evaluación del Área Central de Inglaterra y la Red de Evaluación de Yorkshire y Humber. Las áreas de interés con las que al momento cuenta la UKES y en las que los miembros intercambian información son las artes, la niñez, los jóvenes, la educación, la regeneración económica, la empresa comunitaria de salud, el desarrollo internacional y el transporte. También ha construido puentes entre las diversas comunidades de evaluación y ha desarrollado fuertes lazos con la Sociedad Europea de Evaluación (European Evaluation Society, EES), integrada por sociedades nacionales. Recientemente se unió también a la International Organization for Cooperation in Evaluation (IOCE). Dentro de su página web, la UKES presenta ofertas de trabajo, licitaciones y cursos disponibles relacionados con esta práctica.

Cada año lleva a cabo una conferencia con el objetivo de conocer si los evaluadores logran tener un papel más influyente en la forma en que se piensan y resuelven los problemas.⁹ Entre los temas de interés tratados se suele abordar la relevancia de la evaluación y su rol de cambio; el balance de las elecciones políticas y metodológicas en la evaluación de los programas públicos; los métodos disponibles y las oportunidades que ofrecen las nuevas tecnologías; el valor de las técnicas cualitativas; el compromiso, entendimiento y uso de la evaluación tanto por el público como por los tomadores de decisiones; entre otras. Un tema de especial atención ha sido la posible contribución de la evaluación a la ética, la equidad y el medio ambiente.

La Conferencia de 2013 se centró en el análisis del arraigo y uso de las evaluaciones.¹⁰ Como requisito para lograrlo consideró que la evaluación esté reconocida dentro de la organización, que los pensamientos críticos y reflexivos sean el sello distintivo de la gestión orientada a resultados y que los actores claves tengan la disposición para escuchar los hallazgos y utilizarlos al

⁹ La conferencia anual de 2014 se llamó “Moviéndonos con los tiempos: política de evaluación, práctica y participación”.

¹⁰ Notas de la conferencia de Marlène Läubli, realizada en mayo de 2013 [www.seval.ch/documents/ukes2013.pdf], fecha de consulta: 5 de diciembre de 2014.

realizar sus planes de acción. En particular, se discutió sobre la aportación de la rendición de cuentas. También se debatió sobre los métodos, destacando que hay una tendencia hacia la gestión basada en los resultados y hacia demostrar la rentabilidad social de la inversión, a la par de un mayor énfasis en el *value for money*. Esto presiona a los evaluadores a producir evidencias sobre la forma en que el evaluado ha contribuido directamente al cambio. Varios ponentes sostuvieron que las evaluaciones están teniendo lugar en contextos cada vez más complejos que incrementan el riesgo y la incertidumbre acerca de los posibles resultados, pero que continúan siendo abordados con métodos simples.

Por último, Reino Unido ha estado a la cabeza de la EES,¹¹ integrada por miembros de 74 países, 27 de éstos europeos; también organiza una conferencia bianual¹² y cuenta con grupos temáticos, foros, talleres y seminarios. Además, es sede de EvalPartners, creada por organismos de Naciones Unidas para promover la evaluación en las organizaciones de la sociedad civil (OSC).

Publicaciones especializadas

Las bases de datos de las bibliotecas de dos de las principales universidades del Reino Unido (Cambridge y Oxford) contienen sólo dos libros publicados recientemente. El primero (Crabbé y Leroy, 2008) aborda el tema de la evaluación en el ámbito de la política medioambiental. El segundo (Greener y Greve, 2014) examina el rol de la evidencia en la toma de decisiones sociales y el alcance de su influencia, además de valorar la forma en la cual los recursos escasos pueden ser usados de la mejor manera, principalmente en tiempos de austeridad, y describe el modo en el que investigadores y políticos trabajan juntos de manera eficaz. Hay también un libro (Forss *et al.*, 2011) que discute los métodos disponibles para evaluar políticas públicas en condiciones de complejidad y presenta casos como ejemplos.

Por otra parte, la Sociedad de Evaluación del Reino Unido tiene una revista llamada *The Evaluator*, que dos o tres veces al año publica artículos de

¹¹ Su idioma oficial es el inglés.

¹² La última se realizó en Dublín en octubre pasado y su tema fue “Evaluation for an equitable society, independence, partnership, participation”.

investigaciones recientes en teoría, métodos y práctica de la evaluación tanto en el Reino Unido como en otras partes del mundo.¹³ Dicha sociedad también ha publicado una *Guía para la buena práctica en evaluación*, la cual pretende exponer de manera sencilla los principios aplicables en cualquier disciplina o contexto. Igualmente ha publicado un *Marco para las capacidades de evaluación* que delinea la conducta adecuada para lograr una evaluación de calidad en la promoción de una cultura de profesionalismo para mejorar su práctica.

Adicionalmente, cuenta con *Evaluation, The International Journal of Theory, Research and Practice* editada por Sage Publication para la EES desde 1995. Se trata de una revista interdisciplinaria con trabajos de evaluación en todos los campos: economía, administración, psicología, antropología, etcétera.

Programas académicos

A diferencia de Estados Unidos, no hay una amplia oferta en posgrados enfocados específicamente en evaluación. Las universidades ofrecen programas que incluyen temas de investigación social o análisis cuantitativos y cualitativos dentro de programas más amplios como los de educación. Lo más recurrente son las maestrías en políticas públicas que dedican un módulo a la evaluación y sus métodos. En el Cuadro 2 se presentan los programas más sobresalientes.

A modo de conclusión, la evaluación en Reino Unido es una actividad que presenta avances importantes que actualmente se continúan fomentando, y que impregna a todos los departamentos del gobierno, aunque su calidad no está suficientemente controlada. La formación de evaluadores no se ha consolidado plenamente ya que se muestra un mayor interés por formar profesionales de las políticas públicas en general. La Sociedad de Evaluación contribuye al debate del tema entre sus profesionales y a la realización de publicaciones, aún limitadas en número.

¹³ En el sitio web de la revista el número que se presenta como muestra es de 2011, y no permite el acceso a su contenido ni a números anteriores.

CUADRO 2
Reino Unido: principales programas académicos por universidad

Programa	Universidad/Institución	Descripción
Maestría en intervención social basada en evidencia y evaluación de políticas	Universidad de Oxford	Incluye un curso dirigido a los interesados en estudiar la evaluación de políticas y la eficacia de las intervenciones en problemas sociales.
Maestría en políticas públicas y dirección	Universidad de Glasgow	Su curso de Evaluación, evidencia y política aborda la evaluación ante los retos de la política pública en el mundo real.
Maestría en investigación de políticas públicas	Universidad de Glasgow	Brinda un curso de Evidencia, evaluación y políticas que abarca contenidos de metodología y estadística.
Maestría en políticas públicas	Universidad de Bristol	Uno de sus principales cursos es Información y evaluación de políticas: métodos de investigación y análisis.
Maestría en política pública	Universidad de Oxford	La maestría incluye un módulo sobre evaluación de políticas, en el cual se demuestra la contribución que la evidencia investigativa puede aportar a la política pública.
Maestría en políticas públicas	Universidad de Edimburgo	La evaluación se considera en el módulo sobre Análisis de políticas. Se incluyen enfoques y métodos analíticos, pero también se reconoce la importancia del contexto político. También en el módulo sobre Práctica de la política pública se incluye un curso sobre Métodos de investigación para la evaluación.
Maestría en políticas públicas	Universidad de Nottingham	Considera a la evaluación como un tema clave en el desarrollo de las políticas públicas e incluye su estudio dentro del módulo de Análisis de políticas.

Fuente: páginas web de las respectivas universidades, revisadas en agosto de 2014.

CAPÍTULO III

Francia

*Myriam Cardozo Brum**

En la búsqueda de un estilo propio de evaluación, mejor adaptado a su cultura que el desarrollado en los países anglosajones, Francia ha atravesado por diversas etapas durante el siglo XX, las que se resumen a continuación para luego analizar los eventos ocurridos en los años recientes.¹

En la década de 1960, se produjeron intentos por introducir modelos americanos de evaluación *a priori* ligados a la racionalización del gasto en el marco del desarrollo del Sistema de Planificación, Programación y Presupuestación (PPBS), que en Francia se traduciría en la Racionalización de las Elecciones Presupuetarias (Rationalisation des Choix Budgetaires, RCB). En una difícil colaboración entre la planificación y las ciencias sociales, se desarrolló un fuerte debate sobre la utilización de indicadores sociales y la subordinación de lo social a lo económico.

A mitad de la década de 1970 se produjo la declinación de la RCB y la aparición de la evaluación *ex post* de políticas públicas, la que se concretó en 36 casos aislados entre 1971 y 1978.

* Doctora en ciencias políticas y sociales por la UNAM. Profesora-investigadora, Departamento de Política y Cultura, División de Ciencias Sociales y Humanidades, UAM-Xochimilco, México.

¹ Para profundizar en su desarrollo, véanse Spenlehauer (1998) para el periodo 1960-1990, Thoenig (2005) en 1970-2000, y Cardozo (2012a) en 1980-2000.

En la primera mitad de la década de 1980 se realizaron 153 evaluaciones a las políticas públicas, especialmente en materia de salud, agricultura, empleo y medioambiente. Fue un periodo de reflexión importante, durante el cual destaca la realización de un primer coloquio sobre el tema en 1983. Esta década estuvo marcada por la presentación de dos informes al gobierno (Deleau en 1986 y Viveret en 1989) que implicaron cambios en la conceptualización y los procesos de evaluación y por la aparición de dispositivos de evaluación participativos y pluralistas, con énfasis en ciencias sociales más que en el cuantitativismo.

En 1990 se institucionalizó el Sistema Nacional de Evaluación de Políticas Públicas, con la creación del Comité Interministerial de Evaluación (Comité Interministériel de l'Évaluation, CIME), el Fondo Nacional para el Desarrollo de la Evaluación (Fond National pour le Développement de l'Évaluation, FNDE) y el Consejo Científico de la Evaluación (Conseil Scientifique de l'Évaluation, CSE) para garantizar su independencia, lo que se tradujo en el apogeo de la *expertise* y la evaluación *ex ante*, coordinada por el Comisariado General del Plan (Commissariat General du Plan). A mediados de la década de 1990, Francia había hecho unas 1 500 evaluaciones (algunas más cercanas a la cuasievaluación o a la auditoría), con diversidad de métodos y en casi todos los campos, tanto para el gobierno como para el Parlamento, las colectividades locales, las agencias públicas autónomas, etcétera. La mayoría fueron voluntarias, sólo algunas exigidas por los fondos de la Unión Europea (UE) o la ley. Sin embargo, la evaluación enfrentaba un fuerte escepticismo respecto de su utilidad.

En 1996 se crean dos oficinas parlamentarias destinadas a la evaluación de políticas públicas y de la legislación (esta última, suprimida en 2001) para efectuar contrapeso al Poder Ejecutivo y en 1998 se formó el Consejo Nacional de Evaluación de las Políticas Públicas (Conseil National d'Évaluation des Politiques Publiques, CNE) que sustituyó al CSE, al que se le criticaba que sus evaluaciones fueran pocas (13 en ocho años) y tardadas. Con esto se volvió a reforzar el papel del Comisariado General del Plan, que quedó a cargo del secretariado del nuevo Consejo, y redujo la duración de las evaluaciones a un máximo de 18 meses. Sus trabajos se centraron en la eficacia y el impacto de la acción pública, no tanto en la eficiencia, disminuyendo la importancia de los datos cuantitativos y recurriendo más al apoyo de las ciencias sociales.

El Parlamento puso en marcha una Misión de la Evaluación y el Control (Mission de l'Évaluation et de Contrôle, MEC) del gasto público en 1999, que

ha realizado evaluaciones importantes (formación profesional, ayuda al empleo, etcétera) con la intención de renovar el proceso presupuestal. Ese mismo año se creó la Sociedad Francesa de Evaluación (Société Française d'Évaluation, SFE), que en el futuro sería clave para el debate sobre el tema.

Francia cierra el siglo con una relativa consolidación de la evaluación con su estilo propio y enmarcada en las prácticas impulsada por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), es decir, con un enfoque que prioriza el acercamiento sociológico, plural y participativo, al tiempo que desecha los intentos por introducir la experimentación americana y el *value for money* inglés; pero también con algunos desencantos provocados por la limitada efectividad de los procesos evaluativos promovidos tanto por el Poder Ejecutivo como por el Legislativo.

Institucionalización de la evaluación

Como se desprende del relato previo, Francia inicia el siglo XXI con un cierto grado de institucionalización de la evaluación, pero sujeta a nuevos cambios institucionales. En 2001 el parlamento votó la Ley Orgánica relativa a las Leyes de Finanzas (Loi Organique relative aux Lois de Finances, LOLF) que refuerza su papel de control y evaluación y modifica la lógica previa de una gestión de programas con énfasis en los medios por otra basada en los objetivos y resultados, que será el nuevo objeto de evaluación. Exige su realización por lo menos cada cinco años y condiciona la permanencia de los programas a la obtención de resultados positivos. En esta labor el Parlamento es asistido por la *Cour de Comptes*.

Por su parte, la Ley 2002-2 introdujo reformas a las instituciones sociales y médico-sociales que incluyeron dos mecanismos de evaluación de la calidad: la autoevaluación, cuyos resultados debían comunicarse cada cinco años, y la evaluación externa, obligatoria cada siete años y en función de la cual se renovarían la autorización de los establecimientos cada 15 años.

Un nuevo informe sobre la evaluación de políticas públicas fue emitido por una delegación del Senado (Bourdin *et al.*, 2004). Informaba que, en el periodo 2000-2002, los trabajos de evaluación incluyeron principalmente los temas de lucha contra el sida y desarrollo rural, seguidos de recursos hídricos, seguridad en

las carreteras, empleo no comercial y juvenil y deportes. Predominan los estudios comparados y los análisis territorializados, realizados en similar proporción tanto por medio de métodos estadísticos, modelos econométricos, análisis financieros y jurídicos, como de análisis sociológicos, encuestas de opinión, monografías y estudios de caso. Con presencia menor se encuentran estudios prospectivos y no se incluyó ningún análisis costo-beneficio.

El informe resaltaba la desigual distribución de las capacidades de evaluación en el país y demandaba el reforzamiento de la institucionalización de dicha función, localizándola en el corazón mismo de la reforma del Estado. Recomendaba procesos más transparentes y participativos, evaluaciones de calidad y utilidad, así como su organización al más alto nivel de autoridad y su control por parte de comisiones en que participaran los sectores más importante de la administración pública y mantuvieran mayor contacto con los solicitantes de la evaluación, sin olvidar tampoco su relación con el proceso presupuestario establecido por la Ley Orgánica relativa a las Leyes de Finanzas.

En 2006, el Comisariado General del Plan fue sustituido por el Centro de Análisis Estratégico (Centre d'Analyse Stratégique) y desapareció el dispositivo impulsado por el CNE al ser reemplazado por la Misión de Evaluación de las Políticas Públicas (Mission d'Évaluation des Politiques Publiques), con lo que el Parlamento se volvió responsable de controlar la acción del gobierno, siempre con la asistencia de la Corte de Cuentas (Cour de Comptes). Un año después, de acuerdo con la solicitud de relocalización de la evaluación efectuada por el Senado en 2004, ésta fue elevada a rango ministerial al crearse un Secretariado de Estado de la Prospectiva y la Evaluación de Políticas Públicas (Prospective et de l'Évaluation des Politiques Publiques) dependiente del primer ministro, situación que duró hasta 2009 en que fue suprimido.

En 2008, la evaluación de políticas públicas fue incorporada a la Constitución, señalando al Parlamento como el responsable de llevarla a cabo (artículo 24 y 51-2), para tal fin puede integrar comisiones de investigación y ser asistido por la Corte de Cuentas (artículo 47-2).

En 2012, en búsqueda de un nuevo modelo de acción pública que concilie solidaridad y cohesión social con respeto a los compromisos presupuestarios se impulsó el programa de Modernización de la Acción Pública, a cargo de un Secretariado General dependiente del primer ministro con la instalación

de un Comité Interministerial de Modernización de la Acción Pública (Comité Interministériel de Modernisation de l'Action Publique, CIMAP), que funcionaría como instancia de decisión y arbitraje en el tema con apoyo de una dirección interministerial.

En diciembre de 2013 el CIMAP presentó un informe, seguido de una conferencia de prensa del presidente de Francia el 14 de enero de 2014. En ellos se resalta fuertemente el objetivo de reducción del gasto presupuestal y la generación de evaluaciones cuantitativas, así como la creación de un Consejo Estratégico del Gasto Público, encargado de la evaluación de políticas públicas, que constituiría un indiscutible punto de inflexión respecto de la tendencia introducida previamente por el Programa de Modernización de la Acción Pública. Al mismo tiempo, el presidente mencionó criterios de eficacia, justicia y ecología para realizar “recortes”, lo que implicaría ampliar los puntos de vista meramente presupuestales.

Las evaluaciones terminadas en los últimos dos años y las aprobadas en octubre de 2014 por el Comité de Evaluación y Control de las Políticas Públicas (Comité d'Évaluation et de Contrôle des Politiques Publiques, CEC) de la Asamblea Nacional incluyen temas variados: económicos (apoyo público a las exportaciones, dispositivos de apoyo a la creación de empresas), de desarrollo social (movilidad social de los jóvenes, desempeño de las políticas sociales en Europa), de salud (lucha contra el tabaquismo y el uso de sustancias ilícitas, apoyo médico del Estado y cobertura universal de la enfermedad, medicina escolar), ambientales (paquete “energía-clima Europa”, lucha contra la contaminación del aire), de educación (oferta y necesidades de formación profesional, mixtura social en la educación nacional), de turismo (política de trato al turista), entre otras. Los métodos utilizados son similares en todos los trabajos y giran en torno a la realización de 20-30 audiciones (entrevistas colectivas) en que los evaluadores escuchan a 50-100 personas. Algunas evaluaciones mencionan además la realización de mesas redondas y la observación en campo, que se suman al análisis documental (Asamblea Nacional, 2014).

En cuanto a la metaevaluación, su nivel de institucionalización es aún limitado y concentrado en investigaciones académicas. Lascoumes (1998) realizó una revisando seis evaluaciones aplicadas con distintos enfoques, y Lajoudie y Toulemonde (2002) efectuaron otra a 11 evaluaciones de la región de Limousin.

En fechas más cercanas, Baslé (2013) menciona sólo dos ejemplos,² uno de ellos francés, vinculado con la evaluación experimental para la generalización de una reforma francesa controvertida. En este caso no habría existido suficiente entendimiento entre los objetivos científicos de los evaluadores y los objetivos políticos de los decisores de manera de alcanzar resultados útiles.

Poco es lo que se ha escrito sobre la utilización instrumental de los resultados de las evaluaciones porque Francia, como señala Trosa (2003), a pesar de los esfuerzos alternados del gobierno y la Asamblea Nacional, se ha orientado más a la investigación, ha defendido una evaluación realizada por expertos independientes del gobierno y ha enfatizado el aprendizaje de los actores implicados en cada caso, en lugar de priorizar la contribución gerencial y cuantitativa al servicio de la decisión interna como los países anglosajones.

Profesionalización de la evaluación

Asociaciones profesionales

En los años transcurridos de este nuevo siglo, la SFE ha sido clave en el debate nacional de la evaluación; además, cuenta con grupos y clubes regionales. En 2008, con motivo del Coloquio organizado por la SFE y su contraparte alemana, la Gesellschaft für Evaluation e.V. (DeGEval), se percibía la tendencia al redescubrimiento de métodos que habían dejado de usarse (experimentación o métodos de evaluación *ex ante* como los multicriterios). Cuatro años después Barbier (2012) publicaría una fuerte crítica a este regreso a métodos promovidos por los grandes financiamientos de ayuda al desarrollo, cuyos problemas éticos y metodológicos habían llevado a su abandono en la década de 1960.

En 2009 la SFE definió su primer proyecto asociativo 2009-2013 con siete ejes prioritarios: la constitución del centro de recursos vía internet, el desarrollo

² Uno sobre la evaluación del Ingreso de Solidaridad Activa en 2013 en Francia. El otro referido a las evaluaciones relativas a las políticas de formación y acompañamiento de los demandantes de empleo realizadas por un equipo internacional con un enfoque cuasi-experimental, en distintos países miembros, a solicitud de la Comisión Europea, en 2010.

de jornadas de estudio, la promoción de grupos relativos a la evaluación de políticas públicas sectoriales, la creación de una revista, el desarrollo de redes y de acciones en torno a los tomadores de decisión, el despliegue de la Carta de la Evaluación de Políticas y Programas Públicos y la construcción de un marco de competencias y de desarrollo para los integrantes de la Sociedad.

En 2010 se destaca la realización de un evento sobre evaluación de impacto que buscó dialogar entre evaluadores, metodólogos y especialistas de distintas corrientes. Se discutieron conceptos, técnicas, enfoques y dificultades para su aplicación. Se presentó tanto la perspectiva histórica de la evaluación de impacto como su estimación con base en la experimentación con grupo de control aleatorio (propio del análisis atributivo), pero también la explicación de impactos basados en la teoría de la acción o las cadenas de hipótesis causales (en el marco de los análisis contributivos), lo que permitió analizar el equilibrio entre la estimación cuantitativa y el análisis explicativo, las preguntas que pueden responderse con cada técnica y los problemas éticos y deontológicos de cada enfoque.

Al siguiente año, los socios de la SFE comenzaron a solicitar insistentemente formación y asistencia, mientras sus autoridades perseguían mayor visibilidad de la Sociedad ante los representantes electos, los tomadores de decisión y el público en general. Para satisfacer estas necesidades, desde 2011 se intensificó el número de eventos organizados como se puede ver en el Cuadro 3.

Los cuatro eventos desarrollados en 2011 tuvieron sintéticamente los siguientes contenidos:

- a) “La evaluación de políticas públicas frente a los desafíos de una sociedad en tensión” (junio-julio) resaltó la poca utilidad de la evaluación en los distintos niveles de decisión frente a tensiones como: dificultades económicas, agravamiento de la desigualdad social, amenazas ambientales y restricciones presupuestales. Se discutió sobre la contribución de la evaluación a la acción pública, se desarrollaron talleres y se llevó a cabo una mesa redonda sobre las perspectivas de la evaluación y foros sobre el rol de los investigadores académicos en el desarrollo de la evaluación y sobre metodologías de evaluación de impacto.

CUADRO 3

Francia: principales eventos realizados anualmente por la SFE 2011-2014

Año	Nombre del evento
2011	La evaluación de políticas públicas frente a los desafíos de una sociedad en tensión. La evaluación en España. Desentrañar la complejidad de las políticas públicas. Evaluación en educación, ¿quién evalúa y sobre qué fundamentos?
2012	El desarrollo experimental de la Corte en materia de evaluación. ¿Qué desafíos futuros? 10 años después de la Ley 2002-2, los desafíos de la evaluación externa.
2013	Creación de la Red Francófona de la Evaluación (RFE) y del Observatorio de Evaluación de las Políticas Públicas. Evaluación del desarrollo: ¿qué perspectivas? Metaevaluación y utilidad de la evaluación en la decisión pública. Concepto de metaevaluación. La evaluación participativa: de la toma en consideración de los públicos al poder de acción ciudadano.
2014	La utilidad social, un nuevo criterio de evaluación. Evaluación y desarrollo económico. Profesionalización de la evaluación de las políticas públicas. Evaluación en las comunidades territoriales. El estado del arte y perspectivas de la evaluación del desarrollo sostenible.

Fuente: Boletines de la SFE.

- b) “La evaluación en España” (septiembre) recogió la preocupación sobre la forma de desarrollar las evaluaciones que relacionan el nivel nacional y el de las comunidades, fuertemente descentralizadas, y de integrar sus resultados a la decisión pública y el presupuesto. Para ello se presentaron los objetivos y el programa de trabajo de la Agencia Española de Evaluación (AEE) y del Ministerio de Economía del mismo país.
- c) “Desentrañar la complejidad de las políticas públicas” (septiembre), evento desarrollado conjuntamente con la Universidad de Avignon en que se admitió, siguiendo a Morin y Le Moigne (1999), que las políticas públicas constituyen una construcción colectiva en cambio continuo que incrementa

su complejidad debido, entre otras, a que éstas se entrecruzan entre sí en diferentes escalas ministeriales y territoriales y se enfrentan sistemas de decisión multidimensionales en el marco de restricciones financieras. La complejidad provoca múltiples dificultades, entre ellas su propia evaluación.

- d) “Evaluación en educación, ¿quién evalúa y sobre qué fundamentos?” (diciembre). Su hipótesis fue que los evaluadores harían evaluaciones diferentes en función de su estatus, postura, formación o experiencia, por lo que era necesario discutir su epistemología y su institucionalización en el campo educativo.

Además, en los grupos y clubes regionales³ se debatieron temas como: la evaluación de la política de empleo y de formación, la utilidad social como nueva dimensión de evaluación de las políticas públicas, la formulación de juicios evaluativos a partir de los datos, la evaluación de la política de ayudas económicas, y los estándares de calidad y deontología, entre otros.

Los eventos efectuados por la SFE en 2012 discutieron los siguientes temas:

- a) El seminario “El desarrollo experimental de la Corte en materia de evaluación”, organizado por la Corte de Cuentas, la SFE y el Círculo de la Reforma del Estado (marzo) mostró la tendencia a “redescubrir” la evaluación experimental en Francia. Se discutieron los aportes de este enfoque a la Corte, las dificultades enfrentadas, así como las nuevas relaciones con el Parlamento a partir del análisis de dos casos (vivienda para los más desfavorecidos y biocarburantes).

³ Éstos son numerosos, a continuación se citan los más relevantes. Grupos: Evaluación y desarrollo sostenible, Evaluación de políticas de empleo y formación profesional, Competencias y deontologías de evaluadores externos e internos, Evaluación en salud, Evaluación de las políticas de acción social, Evaluación de políticas culturales, Evaluación en educación, Evaluación y eficacia del desarrollo, Evaluación y utilidad social, de Estándares de calidad y deontología, entre otros. Redes: de Investigación en evaluación de políticas públicas, de Comunidades urbanas, de Comunidades del Sur, etcétera. Clubes: Île de France, Midi-Pirénées, Provence Alpes Côte d’Azur, Bretagne/Pays de la Loire, Rhône-Alpes, etcétera.

- b) Como consecuencia del balance realizado a medio camino al primer proyecto asociativo 2009-2013, la SFE definió temas a desarrollar a partir de 2012 y otros más lejanos, discutidos en el evento “¿Qué desafíos futuros?”. Entre los primeros se decidió la preparación de una guía metodológica; entre los segundos, la instalación de un comité de ética y deontología, la oferta de formación en evaluación, la relación entre economía y acción social y la posibilidad de organizar una conferencia nacional anual.
- c) Es de destacar la jornada nacional “10 años después de la Ley 2002-2, los desafíos de la evaluación externa”, que trató de reunir a todas las partes implicadas en los procesos evaluativos (profesionales, evaluadores externos, autoridades públicas, representantes de los usuarios, etcétera) e intercambiar puntos de vista. Se partió de la idea que la evaluación ya estaba institucionalizada, especialmente en su modalidad de evaluación interna. Se discutió sobre las tensiones persistentes entre diferentes concepciones de la evaluación, el lugar que le corresponde a cada uno de los involucrados (especialmente los usuarios), los términos de referencia (TR) y los principios metodológicos, entre otros.

Finalmente, la Red de Investigadores en Evaluación de Políticas Públicas de la SFE organizó un ciclo de seminarios sobre los desafíos de la evaluación participativa (junio) y otro sobre los métodos de síntesis de conocimientos (diciembre).

En 2013 se incrementó el número de eventos según el siguiente detalle:

- a) Se creó la Réseau Francophone d'Évaluation (RFE) y un Observatorio de Evaluación de las Políticas Públicas y se llevaron a cabo diversos eventos que muestran las preocupaciones de la SFE: las nuevas perspectivas de la evaluación, la metaevaluación y su utilidad para la decisión pública, la discusión del balance del proyecto asociativo 2009-2013 y la definición del siguiente para 2014-2019, el análisis del contexto externo y de la propia SFE, la evaluación participativa, en paralelo a la reunión de clubes, grupos, redes, y publicaciones diversas.
- b) Se realizó la jornada de estudio “Evaluación del desarrollo: ¿qué perspectivas?” (febrero), dedicada a analizar el compromiso asumido por Francia en

el marco de la agenda sobre la eficacia de la ayuda internacional, conocida como Declaración de París, la evaluación de las intervenciones realizadas, la apropiación por las partes involucradas, y la promoción del desarrollo de la capacidad de evaluar por parte de los países receptores a partir de experiencias concretas recientes.

- c) Se complementó el ciclo de seminarios iniciado en 2012 por la Red de Investigadores en Evaluación de Políticas Públicas de la SFE con la discusión sobre la institucionalización de la evaluación y la decisión política (marzo). De particular interés resulta el tema elegido en abril para su cuarto seminario de investigación “Metaevaluación y utilidad de la evaluación en la decisión pública”. En la introducción se resalta la ausencia de un concepto de metaevaluación que goce del consenso de los teóricos, la definición más simple de Scriven (1969), la posibilidad de aplicarse a una o varias evaluaciones de un tema, proyecto o política y los elementos involucrados: el análisis de los resultados obtenidos, la calidad de las prácticas, las lógicas teóricas, los métodos, la formación de los juicios de valor, así como su utilidad para los decisores y la sociedad en general, es decir que incluyó la validez científica pero también tomó en cuenta la finalidades sociales. El seminario también fue dedicado a la discusión teórica del concepto de utilización de los resultados, a la posición de sus principales autores, y a la presentación de tres ejemplos.
- d) El tema de la utilización de resultados fue retomado en París, también en abril, con motivo de un Seminario de Investigación sobre el Concepto de Metaevaluación y en las jornadas organizadas por la SFE en junio en Grenoble.⁴
- e) El grupo de Evaluación Participativa organizó la jornada de estudio “La evaluación participativa: de la toma en consideración de los públicos al poder de acción ciudadano” (diciembre). Este ejercicio buscó contribuir a reforzar la democracia preguntándose sobre la posibilidad real de este tipo de evaluación, las condiciones y medios que permiten su realización, en especial, la posibilidad de que los ciudadanos coordinen el proceso de cuestionamiento y de comprensión de la lógica y efectos de la acción pública.

⁴ Seminario de investigación organizado por la Red de Investigadores en Evaluación de Políticas Públicas (RCEPP-SFE), París, 18 de abril de 2013 y Jornadas Francesas de Evaluación, Grenoble, 6 y 7 de junio.

El 2014 comenzó con mucho trabajo y algunas inquietudes. El trabajo se relaciona con el nuevo proyecto asociativo que se votaría en junio, al que se adicionan temas controvertidos como la eventual acreditación de evaluadores, y otros como la aplicación de una política de comunicación más sistemática que refuerce las relaciones con todos los actores o la preparación de un barómetro anual de la evaluación que permita capitalizar los trabajos del Observatorio y convertir a la SFE en una referencia obligada en el tema. En cuanto a las inquietudes, éstas nacieron por la falta de claridad respecto de los criterios que guiarán las evaluaciones en el ya mencionado informe presentado en diciembre de 2013 por el CIMAP y la conferencia de prensa del presidente de Francia el 14 de enero de 2014. En ese sentido la SFE acepta que el criterio de eficiencia, presente en todas las guías y manuales, debe ser integrado en el análisis pero no está de acuerdo en realizar exclusivamente evaluaciones cuantitativas.⁵

Este año se realizaron los eventos mencionados a continuación:

- a) Entre las jornadas de estudio, se llevaron a cabo dos (junio y noviembre) dirigidas a “La utilidad social, un nuevo criterio de evaluación”. En éstas se propuso a la utilidad social como herramienta de apreciación y juicio valorativo en las prácticas evaluativas, junto con los criterios relativos al desempeño de la gestión que son los dominantes (coherencia, pertinencia, efectividad, eficacia y eficiencia). A diferencia de estos criterios, la utilidad social no surge de un registro factual o contrafactual sino de la controversia entre los actores, lo que supone una cultura evaluativa de tipo cualitativo, pluralista, democrática, participativa e interactiva. Si bien este nuevo criterio no está institucionalizado, se usa y se reconoce cada vez más por diversos sistemas de actores.
- b) Otra jornada relevante, con participación tanto de investigadores como de funcionarios de diferentes niveles de gobierno, se abocó al tema “Evaluación y desarrollo económico” (julio) para analizar los desafíos y márgenes de maniobra de la intervención pública en dicho campo, las dificultades existentes para evaluar políticas económicas y las orientaciones estratégicas impuestas a los fondos europeos y sus evaluaciones.

⁵ Si bien los indicadores, las medidas y las cifras son necesarias, la SFE considera que no agotan las preguntas de evaluación.

- c) Por su parte los grupos y clubes trabajaron en temas como: la evaluación filmada, la utilización de historias de vida en la evaluación, los métodos de análisis multicriterios, la observación participante, la revisión crítica de indicadores, la profesionalización en la evaluación, la evaluabilidad de una política cultural, el análisis de la política ambiental, entre otros. El Grupo de Evaluación de Políticas y de Empleo y de Formación Profesional retomó dos temas relevantes en su reunión: evaluación de impacto y metaevaluación. Al mismo tiempo el Club de Île de France dedicó una sesión de trabajo a la profesionalización en la evaluación de las políticas públicas.
- d) A fines de 2014 se llevaron a cabo diversos eventos relativos a: discusión del informe “Las múltiples escenas de la evaluación”, y las jornadas nacionales “Evaluación en las comunidades territoriales” y “El estado del arte y perspectivas de la evaluación del desarrollo sostenible”. La SFE organizó también tres talleres sobre los principales desafíos territoriales: la infraestructura, las políticas migratorias y las políticas sociales. Para fin de año se programó contar con el primer barómetro de las aproximadamente 1 077 evaluaciones del periodo 2007-2013. El mismo incluirá: a) análisis cuantitativos de datos estadísticos sobre el estado de la evaluación en Francia en 2013; b) análisis cualitativos de miembros de la SFE y de entrevistas con informantes-clave; y c) enfoque cuanti-cualitativo de ciertas evaluaciones, de resultados importantes de renovación pública, etcétera.
- e) Posteriormente a la aprobación del nuevo proyecto asociativo 2014-2017, también se estableció el programa de formación demandado por los miembros de la SFE para 2015 compuesto por cursos de iniciación en evaluación, elaboración de TR, métodos y técnicas, participación de los ciudadanos y resultados.

La multiplicidad de eventos y temas referidos muestra la fuerza adquirida por la reflexión y el debate en el contexto de la SFE y, en general, en Francia.

Publicaciones especializadas

Francia ha publicado numerosos libros y no tantos artículos, probablemente debido a que carece de una revista especializada en evaluación, uno de los

proyectos ya mencionados de la SFE. En consecuencia sólo se mencionan algunos de los libros y artículos más recientes.⁶

En 2008 se publicaron tres libros relevantes sobre el tema (de Perret; Vollet y Hadjab; y Baslé), en el siguiente año destacan dos publicaciones (Fouquet y Méasson; y un artículo coletivo en *Revue Cité*), y en 2010 se produjeron seis adicionales (Association France Qualité Publique; un artículo coletivo en *Revue Française de Socio-Economie*; Bozio y Grenet, coord.; Mouterde y Trosa; Braconier y Cauquil; y Offredi y Ravoux).

Destaca la publicación dedicada por Mouterde y Trosa al análisis de los 20 años de evaluación transcurridos en Francia (desde el informe Viveret de 1989) y 10 desde la creación de la SFE. La misma señala los avances logrados y resalta los desafíos pendientes: desarrollo de una verdadera cultura de la evaluación, más amplia difusión pública de la misma, profundización de los conocimientos, mayor acercamiento entre disciplinas y métodos, focalización estratégica de los trabajos, entre otros. Adicionalmente, la *Revue Française des Affaires Sociales* dedicó dos números a la evaluación de políticas públicas, sanitarias y sociales y se publicaron otros artículos en *Pouvoirs Locaux*, *Esprit*, *Informations Sociales*, *Regards sur l'Actualité* y *Perspective Gestions Publiques*.

La SFE también publica Cuadernos; en 2011 apareció el número 6 dedicado a “Evaluación de los impactos de los programas y servicios públicos”. Incluye una revisión histórica desde la década de 1920, presenta una decena de métodos ilustrados analizando su utilidad y aplicabilidad, así como estudios de casos preparados por el Grupo de Estándares de Calidad y Deontología, tanto con impactos esperados como no esperados. También publicó otro relativo a las prácticas de evaluación de políticas públicas que sintetiza los resultados de los talleres desarrollados en 2010, incluyendo la ayuda social a la infancia, las experiencias en salud, así como una guía de análisis de evaluaciones. Ese año fueron publicados dos artículos en las revistas *Gestion et Finances Publiques* y *Pouvoirs Locaux*.

⁶ Una bibliografía comentada y actualizada a julio de 2014 fue publicada por la École Nationale d'Administration (ENA) [www.ena.fr/index.php?fr/content/download/2529/16807/file/bib_evaluation_politiques_publicques_sf.pdf], fecha de consulta: 21 de marzo de 2014.

Barbier (2012) publicó un extenso e interesante artículo en que criticaba el enfoque experimental y planteaba la necesidad de regresar a las fuentes intelectuales de la evaluación de políticas públicas, al pluralismo metodológico y a la epistemología de las ciencias sociales con sus distintas concepciones sobre la causalidad. Consideraba que otros enfoques se estaban reposicionando entre funcionarios y evaluadores y preveía que el futuro de la evaluación en Francia no continuaría por el camino experimental. Ese mismo año aparecieron otros dos libros de importancia (Baron y Matyjasik, 2012; y Battesti, 2012),

El siguiente año cerró con la publicación de un nuevo libro (Bouquet, 2013) y diversos trabajos: “Évaluer une action publique”, pequeña guía dirigida a representantes y otros decisores; dos artículos incluidos en la revista *Sociologies Pratiques* y otros dos en *Pouvoirs Locaux* y la *Revue de l’IRES*; y el Cuaderno núm. 8, titulado “L’évaluation pour l’efficacité de la coopération au développement”, que recupera los trabajos realizados en 2008-2010 sobre la contribución de la evaluación a la eficacia de la ayuda, los tipos específicos de evaluación conjunta y el análisis del impacto en términos de su utilidad y métodos considerados. También la idea expresada por el nuevo presidente de la SFE en su primer Boletín de Información, en el sentido que la evaluación goza de consenso general pero no responde plenamente a lo que de ella se espera en relación con su contenido y fines operativos, por lo que resulta fundamental compartir experiencias e incentivar su transversalidad, sin descuidar la profundidad de los casos sectoriales y recoger y sintetizar propuestas para el nuevo proyecto asociativo 2014-2019.

En 2014 la *Revue Française d’Administration Publique* ha dedicado su último número al estado del arte y las controversias en torno a la evaluación de las políticas públicas. Incluye una introducción de la presidenta de la SFE y diversos artículos que tratan, entre otros, los temas de institucionalización de la evaluación, indicadores, criterios, control de gestión, metaevaluación y calidad de las evaluaciones.⁷ En el siguiente número, la revista programa incluir los estudios de impacto. Además, ese año apareció otra publicación importante (Aberlen, 2014) que presenta los resultados de un seminario que gira en torno a la política

⁷ Fueron publicados en la *Revue Française d’Administration Publique* (2014), núm. 148, dedicada a “La evaluación de las políticas públicas: estado del arte y controversias”.

de apoyo internacional, especialmente promovida luego de la Declaración de París de 2005 y un artículo en la revista *Politiques et Management Public*.

También existen documentos conceptuales y metodológicos sobre metaevaluación en algunas páginas electrónica⁸ pero no se localizan informes de metaevaluaciones realizados recientemente.

Programas académicos

En 2004 se creó la Maestría en evaluación y seguimiento de las políticas públicas del Instituto de Estudios Políticos y la Universidad Lumière, Lyon, la más específica en el contexto francés. Este año, para festejar su primera década, organizó un coloquio, conjuntamente con el Club de Rhône-Alpes sobre la gobernanza o gobernanzas necesarias para la evaluación.

En el Cuadro 4 se sintetiza el conjunto de programas franceses de posgrado relacionados, directa o indirectamente, con el tema.

Mientras algunos de los programas presentados se focalizan claramente en la evaluación de políticas públicas y sociales, otros enfatizan la evaluación de proyectos, tanto públicos como privados.

En síntesis, se considera que la evaluación ha arraigado en Francia, a pesar de su cambiante proceso de institucionalización, pero se perciben recientes titubeos respecto del enfoque propio que ha venido defendiendo como alternativa al modelo anglosajón. Muestra aún un déficit de programas universitarios y de publicaciones periódicas especializadas en la materia. Definitivamente, su mayor riqueza consiste en la intensa actividad de intercambio y discusión que se produce en el seno de la Sociedad Francesa de Evaluación.

⁸ Por ejemplo “Méta-évaluation et évaluation stratégique” [<http://controverses.sciences-po.fr/archive/aidedeveloppement/4.html>], fecha de consulta: 30 de noviembre de 2014.

CUADRO 4

Francia: oferta universitaria de formación en evaluación

Programa	Universidad/Institución	Descripción
Doctorado y maestría en políticas públicas y desarrollo	Escuela de Economía de París	Responde al déficit de formación en evaluación de la acción pública tanto de los países desarrollados como en vías de desarrollo.
Maestría en evaluación y seguimiento de las políticas públicas	Instituto de Estudios Políticos y la Universidad Lumière, Lyon	Cubre todo el proceso evaluativo con la revisión conceptual, teórica y metodológica y la realización de una estancia aplicada de cinco meses de duración.
Maestría en administración y evaluación de políticas sociales	Universidad Pierre Mendès, Grenoble	Forma responsables de políticas públicas y su evaluación en campos sociales, médico-sociales y sanitarios frente a los desafíos económicos y políticos contemporáneos.
Maestría en ciencias políticas	Universidad de París 13	Desarrolla una especialidad para conducir y evaluar las políticas públicas mediante un enfoque pluridisciplinario en diferentes sectores: social, urbano, empleo, medioambiente, etcétera.
Maestría en ciencias políticas	Universidad Julio Verne, Picardía	Ofrece una especialidad en evaluación de políticas públicas, que incluye métodos y técnicas cuantitativas y cualitativas dirigidas a sectores sociales, salud, empleo, cultura, transporte, urbanismo, etcétera.
Maestría en economía y gestión pública	Universidad de Rennes	Imparte formación en economía y evaluación de políticas y programas públicos con especialidades en desarrollo local y sostenible, análisis de proyectos, etcétera.
Maestría en economía aplicada	Universidad Aix Marseille	Enseña herramientas de evaluación económica y técnicas de gestión. Cuenta con una especialidad en Administración y evaluación de proyectos y programas públicos.
Maestría en evaluación y gestión de proyectos	Universidad de París 1	Prepara especialistas tanto para organismos públicos como para empresas, incluyendo las diferentes dimensiones de la evaluación económica de las decisiones y el tratamiento de datos, tanto a corto como largo plazo.
Maestría en análisis y evaluación de proyectos	Universidad de Rennes y Universidad de Yaounde (Camerún)	Refuerza las capacidades africanas en este campo con contenidos de evaluación de políticas y programas. También ofrece a sus candidatos una estancia en calidad de consejeros o decisores en la administración pública o privada.

Fuente: páginas web de cada institución.

CAPÍTULO IV

Canadá

Myriam Cardozo Brum

Canadá ha sido considerado un país pionero en materia de evaluación (Saint Clair, 1993), con similar grado de institucionalización y madurez que en Francia, pero con un objetivo sensiblemente diferente (Jacob, 2006). Su consolidación es fruto de un largo proceso que se resume a continuación hasta aproximadamente el año 2000.¹ A fines de la década de 1960, el Secretariado del Consejo del Tesoro introdujo el Sistema de planificación, programación y presupuestación (PPBS), ligado a la evaluación presupuestaria *a priori*. Sin embargo, aunque existen antecedentes previos, el monopolio de la evaluación por parte del Poder Ejecutivo Federal dio inicio en 1977 con la emisión de la Circular de la Oficina del Contralor General de Canadá “Evaluación de programas efectuada por los ministerios y organismos”, aplicada hasta 1995. Dicha Oficina coordinaba la formación de evaluadores, elaboraba las guías y demás documentos de apoyo, definía estándares metodológicos, se guiaba por los criterios de eficacia, economía y eficiencia, y participaba en la selección de los equipos responsables de realizar las evaluaciones, concebidas como una ayuda a la gestión y toma de decisiones. En su apoyo se publicó en 1981 una “Guía sobre la función de la evaluación de programas” y los “Principios para la evaluación de programas por los ministerios y organismos federales”.

¹ Para una revisión detallada de este periodo, véanse Steve (2006), Bannon y otros (2009) y Cardozo (2012a)

Al inicio de la década de 1980, en el contexto de la crisis económica, las evaluaciones utilizaron métodos costo-beneficio, costo-eficacia o *value for money* que sirvieron para brindar información al Parlamento con objeto de reducir el gasto público mediante la supresión de programas. En 1987 cada ministerio contaba con una unidad encargada de evaluar todos los programas por lo menos una vez cada cinco años, alcanzando entre 300 y 400 evaluaciones anuales.

En 1991 la Oficina del Contralor General preparó un nuevo plan llamado “Abordar los noventa: perspectivas gubernamentales para la evaluación de programas” por la que buscaba incrementar la productividad de la evaluación. Unas 60 instituciones de gobierno, entre éstas todos los departamentos grandes y medianos, contaban con una unidad de evaluación corporativa. Sin embargo Lacasse (1993), al realizar una primera metaevaluación, caracterizó este proceso como un “rito” manipulable que intentaba evaluarlo todo pero no incursionaba en sus aspectos más relevantes (pertinencia, vinculación interministerial, recomendaciones de cambios o incluso de cancelaciones, etcétera) ni en la utilización de sus resultados. Ese mismo año se creó la Sociedad Canadiense de Evaluación (Canadian Evaluation Society, CES).

En 1994 comenzó a aplicarse una política de Examen de los Programas para desarrollar una gestión por resultados con tres objetivos: *a*) verificar si el programa es congruente con las necesidades que intenta cubrir y con las prioridades del ministerio en cuestión; *b*) si con el presupuesto disponible logra sus objetivos sin imponer restricciones no deseadas; y *c*) si utiliza los medios más apropiados y eficientes para alcanzarlos. De esta manera se buscaba recuperar la confianza de los ciudadanos y reducir el gasto gubernamental, cambiando el enfoque estadístico y contable previo por otro enfocado hacia el logro de los objetivos y metas. Se involucraron los grupos de interés (gobernanza) y se redujeron costos debido al cierre o modificación de los programas, pero 60% de las evaluaciones continuaron concentrándose sólo en los proceso de implementación y careciendo de análisis de su pertinencia y de medición de impacto. Las mismas continuaban en manos del Poder Ejecutivo ante un Legislativo completamente ausente (Tourgeon, 1998).

Institucionalización de la evaluación

En el 2000 entró en vigencia la Ley de la Administración Pública (Loi sur l'Administration Publique, LAP) que orienta la gestión gubernamental por resultados, lo que promueve la realización de evaluaciones *a posteriori*, con amplia participación de los funcionarios (considerados sus beneficiarios directos) y centrada en su aprendizaje sobre el uso de las evaluaciones para la toma de decisiones, la rendición de cuentas obligatoria a los ciudadanos y el reforzamiento del Poder Legislativo en la valoración de la gestión.

Para su aplicación el gobierno clarificó el concepto de evaluación y previó su aplicación en torno a los criterios de pertinencia, eficacia, eficiencia, subsidiariedad e impacto, respetando también la validez, fidelidad a la situación, credibilidad y utilidad. Las administraciones deben nombrar un jefe y un comité de evaluación, cuyas responsabilidades y normas también fueron establecidas en el *Cadre de gestion et de responsabilisation axé sur les résultats*. Sin embargo, la LAP no obliga a hacer evaluaciones, sólo faculta al Consejo del Tesoro a exigir las a cualquier ministerio u organismo.

En consecuencia, los ministerios, los organismos o el propio Consejo del Tesoro deciden y realizan las evaluaciones, caracterizadas por ser internas, aunque también se hacen evaluaciones interministeriales y del gobierno en su conjunto. Adicionalmente las impulsan el Ministerio de Finanzas y el Parlamento. A pesar de lo estipulado por la LAP, la mayoría no se hace pública, por lo que no contribuye a la transparencia.

En 2003 se revaloró la función del contralor general, ubicándolo junto al Secretariado del Consejo del Tesoro y, en 2006, se expidió la Ley sobre la Responsabilidad; ambos muestran que la cultura de la evaluación se ha difundido a nivel federal y que la rendición de cuentas se presenta como una prioridad del gobierno (Jacob, 2006).

Un año después, el Centro de Excelencia en Evaluación del Secretariado del Consejo del Tesoro realizó un “Examen de la calidad de las evaluaciones en los ministerios y los organismos” que incluyó una centena de posteriores evaluaciones a la introducción de cambios al inicio del nuevo siglo. En general, la calidad resultó aceptable y los informes bien escritos, salvo 23% que fueron juzgados inadecuados. En la mayoría de las evaluaciones hay una concentración en el logro de éxito (94%), pertinencia (74%) e implementación (72%); mientras

los temas de rentabilidad (44%) y de las prácticas de gestión (47%) tienen poca presencia. Las mayores debilidades se concentran en la falta de descripción de los enfoques y metodologías aplicadas (46%) y de los elementos y preguntas de investigación (35%); por el contrario, las mayores fortalezas aparecen en torno a la descripción del programa (49%), metodología adecuada (42%), aplicación de recursos (40%) y pertinencia del programa a las necesidades (40%).

También en 2004, a raíz de la preocupación por la calidad de las evaluaciones federales, el Secretariado del Consejo del Tesoro realizó una metaevaluación de considerable envergadura en torno de su pertinencia, metodología y credibilidad de sus conclusiones. Si bien se identificaron algunas debilidades, Daigneault (2010) considera que, en general, las conclusiones de las evaluaciones son pertinentes y su nivel de credibilidad aceptable.

En 2007 se modificó el Sistema de Administración del Gasto y en 2009 se revisó la política de evaluación de forma que coadyuvara con el primero al enfatizar la evaluación del gasto junto con la relevancia y desempeño de los programas, con la expectativa de lograr su pleno cumplimiento por parte de los departamentos para alcanzar una mejor utilización en la toma de decisiones y una gobernanza más eficaz.

Lahey (2010) describe las ocho características clave del sistema:

1. Énfasis tanto en monitoreo como en evaluación.
2. Unidades de evaluación interna en la mayoría de los departamentos federales, con líderes centrales.
3. Fundamentos bien definidos del conjunto de roles y expectativas de la evaluación: políticas, estándares, guías.
4. Contrapesos para apoyar la neutralidad/independencia del evaluador.
5. Mecanismos de vigilancia para reforzar la credibilidad y ofrecer control de calidad.
6. Flexibilidad y voluntad para aprender y hacer ajustes.
7. Transparencia como un valor subyacente del sistema.
8. Compromiso permanente con la construcción de capacidades.

A su vez, sintetiza los resultados obtenidos por este sistema como:

- Una fuerte inversión de recursos en evaluación.
- Cobertura del 20% del gasto directo en programas de gobierno (unas 230 evaluaciones por año) y expectativa de lograrlo en su totalidad.
- Amplio uso de los resultados de las evaluaciones.
- Utilización de la información de Monitoreo y Evaluación (M&E) para tomar decisiones de financiamiento.

Finalmente plantea las lecciones que nos deja la experiencia canadiense:

- Un sistema así requiere de muchos recursos de todo tipo (dinero, líderes, etcétera) por lo que se justifica si se necesita y se usa.
- Se requiere llevar a cabo el sistema a largo plazo como una parte natural del funcionamiento institucional.
- Un elemento crítico para mantener el sistema es contar con personal calificado, de ahí la necesidad de construir capacidades, asegurada entre otras por la certificación hecha por la Sociedad Canadiense de Evaluación.

Concluye que cada país tiene su propio contexto, por lo que estos aprendizajes deben ser adaptados, pero se debe tener claro que el camino es largo.

Por último, dentro del tema “arte y cultura” es posible consultar los informes de las evaluaciones realizadas en el periodo 2000-2014 en el sitio del Government of Canada. En el último de estos años se mencionan las referidas a los fondos dirigidos al arte, a las industrias culturales, del programa del Grupo Fiduciario de los Edificios de los Padres de la Confederación y de los fondos para la formación en el sector de las artes.

En el contexto canadiense es interesante analizar el caso de la provincia de Quebec, donde desde 1989 se fundó el Grupo de Responsables en Análisis y Evaluación de Programas (GRAEP), con la participación de funcionarios, para promover el desarrollo de la evaluación de programas en el gobierno de Quebec. En su informe publicado en 2005 *Cinq années de gestion axée sur les résultats au gouvernement de Québec*, dicho gobierno indica una mejoría en el desarrollo de una verdadera cultura de la evaluación, que debe estar integrada al ciclo de gestión de los programas.

En 2009 el GRAEP publicó un manual (Bannon *et al.*) dirigido a los propios funcionarios y también a los evaluadores, gestores, y decisores en general, con el objetivo de informar sobre el contexto y la práctica de la evaluación en Quebec, así como los métodos disponibles y los ejemplos de documentos a elaborar (convocatorias, TR, etcétera). Además, el GRAEP organiza anualmente un Premio de Excelencia en Evaluación.

Algunos ministerios y organismos de Quebec cuentan con referencias específicas sobre los objetos y las solicitudes de evaluación, sus aprobaciones, el rol del comité coordinador de la evaluación, la difusión de su informe y el seguimiento de sus recomendaciones. Además existe una Ley de acceso a la información (*Loi sur l'accès aux documents des organismes publics et sur la protection de renseignements personnels*) que debe ser respetada por funcionarios y evaluadores, garantizando una mayor difusión del proceso y de los resultados de evaluación que en el resto del país. Las evaluaciones interministeriales han sido pocas hasta años recientes, pero han empezado a surgir a pesar de su complejidad.

Internacionalmente, las evaluaciones canadienses son reconocidas; una de ellas referida a un programa de salud dirigido a las “Primeras naciones” acaba de ser elegida como una de las nueve experiencias mundiales de evaluación dignas de contarse por su contribución a la mejora de las condiciones de vida de sus beneficiarios en el concurso “Evaluations that make a difference” (Eval Partners, 2013).

Trosa (2003) ha caracterizado sintéticamente al sistema canadiense de la siguiente manera: existe una cultura nacional de la evaluación que se encuentra integrada en la política de modernización de la gestión, la promoción de la evaluación ha dependido del Ministerio de Finanzas y la Corte de Cuentas (el Parlamento ha tenido menos peso), su objetivo es realizar estudios científicos sobre la pertinencia de los programas, lo que constituye un requisito administrativo bien definido, la evaluación puede realizarse en forma interna o externa pero siempre conducida por unos TR claros y precisos que permiten el desarrollo de una evaluación independiente, y la calidad de los trabajos está asegurada por la Corte de Cuentas y el Tesoro.

Se debe tener claro que la evaluación en Canadá es un mecanismo interno al gobierno, muy diferente de la evaluación democrática, participativa y pluralista francesa, por lo que garantizar su independencia resulta relevante. Si bien los dos sistemas intentan aumentar la responsabilización y mejorar la calidad de

los servicios públicos, sus caminos son diferentes ya que Canadá no transparenta los resultados de las evaluaciones ni invita a los ciudadanos a participar en éstas, lo que le ha permitido el desarrollo de la evaluación sólo a nivel gubernamental. La aplicación de evaluaciones se realiza con flexibilidad y mediante persuasión ya que no existen sanciones por incumplimiento en la aplicación de políticas, lo que provoca avances desiguales, pero, en conjunto, la estrategia está cumpliendo sus objetivos, sin generar impactos no deseados. La obligación de evaluar se funda en la sensibilidad de los funcionarios por lo que sus alcances y calidad son variables (Jacob, 2006).

También existe una crítica permanente al dispositivo de evaluación canadiense, entre otras razones, porque a veces éste se limita a la aplicación de una encuesta. Pero esto no significa un rechazo al mismo, definitivo ni temporal, sino la necesidad de su adaptación a las expectativas que suscita, en una búsqueda continua por mejorar su calidad.

Profesionalización de la evaluación

Asociaciones profesionales

La CES es una asociación bilingüe (inglés y francés), sin fines lucrativos y dedicada al desarrollo de la teoría y la práctica de la evaluación, que cuenta con el apoyo del gobierno de Canadá. En 2003 estaba conformada por 11 secciones regionales que representaban a todo el país y por 1 922 miembros provenientes del gobierno, el sector privado, la educación y organizaciones sin fines de lucro, con perfiles académicos y disciplinarios variados. Organiza una conferencia anual, entrega el premio a la mejor tesis en el tema y tiene su propio código de ética. Publica la lista de los evaluadores acreditados por región (actualmente 256 en total) y cuenta con una amplia base de datos (Grey Literature).

Durante su conferencia anual en 2003 llevó a cabo un taller de funcionarios, universitarios y consultores para analizar mediante una matriz de Fortalezas, Oportunidades, Debilidades y Amenazas (FODA) el estado de la evaluación en el país. Dicho ejercicio arrojó que la principal fortaleza es el dominio metodológico de las técnicas de recolección y análisis de datos, mientras que el pragmatismo de los evaluadores (que son observadores que ofrecen información

pertinente, pero no teóricos) fue considerado una debilidad porque se consideró que se limitan a medir el rendimiento o a hacer un mero seguimiento al programa. El problema más grave es que la evaluación aparezca desconectada de las necesidades concretas de la gestión para la que es contratada y que no se dé seguimiento a las recomendaciones.

En 2005, la conferencia anual de la CES se realizó conjuntamente con la AEA, recibiendo más de 1 200 ponencias provenientes de 43 países, especialmente de sociedades europeas y de la australiana. La última conferencia realizada en Ottawa, en junio de 2014, festejó las contribuciones efectuadas por la Sociedad a lo largo de 35 años.

La Sociedad de Evaluación de Programas de Québec (Société Québécoise d'Évaluation de Programme, SQEP), uno de los capítulos más fuertes de la CES, adhiere al Código de Ética de la última sobrecompetencia, integridad e imputabilidad. Adicionalmente, utiliza las normas publicadas en 1981 y revisadas en 1994 por el Joint Committee on Standards referidas a la utilidad, rectitud, factibilidad y exactitud. También aplica las reglas instituidas en 2005 por el United Nations Evaluation Group (UNEG): *Règles d'évaluation applicables au sein des systèmes de Nations Unies*. Actualmente, se encuentra organizando su vigésimo tercer coloquio sobre análisis críticos de los aportes de la evaluación a las conductas de cambio.

Publicaciones especializadas

La CES, junto con la Universidad de Calgary, cumple desde 1986 la importante labor de publicar el *Canadian Journal of Program Evaluation*, también bilingüe (inglés y francés), que incluye tres números por volumen anual.

Actualmente, acaba de publicarse el número 1 del volumen 29, correspondiente a la primavera de 2014, que incluye artículos sobre teoría y práctica de la evaluación, metodología, estándares, utilización de los trabajos, metaevaluación, reseña de libros y especialmente, resultados de investigaciones empíricas originales. En materia de metaevaluación se han publicado: 1) un artículo de Hum y Simpson (2002), que analiza la evaluación de los programas públicos de capacitación; 2) otro de Patton (2013) sobre la evaluación de la Declaración

de París, revisando su diseño, aplicación, recolección de datos y conclusiones; y 3) uno más de Jacob y Desautels (2014) que analiza una muestra no probabilísticas de 20 evaluaciones de programas para población autótona.

Por otra parte, el Centre de Recherche et d'Expertise en Evaluation de la École Nationale d'Administration Publique (ENAP) publica la *Revue Internationale en Évaluation du Développement*, en forma de un boletín electrónico mensual que dio inicio en mayo del 2013 y que se difunde en África, América y Europa. Contiene sobre todo noticias sobre eventos, creación de sociedades de evaluación, nuevas publicaciones, entre otros.

Además de estas publicaciones periódicas, sólo se ha identificado un reducido número de artículos sobre evaluación publicados en revistas de administración pública, servicio social, educación o negocios (por ejemplo, Ridde *et al.*, 2003; Howlet, 2009; Le Ster, 2011; o la *Revue Mesure et Évaluation en Éducation*, entre otros) y algunas páginas electrónicas con aspectos conceptuales y metodológicos de la metaevaluación.² La publicación de libros es escasa, sobresalen los de Ridde y Dagenais (2009) y Mc David *et al.* (2013).

Programas académicos

En materia de formación en evaluación, en el año 2000 Canadá contaba con 25 instituciones que ofrecían más de 100 cursos vinculados a programas de ciencia política, administración pública, economía, entre otros. Este desarrollo profesional también es promovido por la CES mediante una amplia oferta de talleres especializados y la difusión de una reconocida “Serie sobre las competencias esenciales” que incluye aspectos conceptuales, diseño y puesta en práctica de la evaluación, obtención de resultados y comunicación de hallazgos.

En Canadá (exceptuando la provincia de Quebec) se imparten, entre otros, los programas académicos que aparecen en el Cuadro 5, a los que pueden agregarse algunos cursos cortos de una o dos semanas de duración.

² Un ejemplo es el de Patricia Rogers, “Meta-evaluation bibliography [<http://www.aral.com.au/resources/meta-eval-bib.html>], fecha de consulta: 10 de septiembre de 2014.

CUADRO 5
*Oferta universitaria de formación en evaluación en Canadá,
 exceptuando la provincia de Quebec*

Programas/cursos	Universidad/empresas	Descripción
Doctorado en filosofía en medición, evaluación y metodología de investigación.	Universidad de British Columbia, Vancouver	Preparación de especialistas para el campo educativo, psicológico, salud y contextos sociales con métodos cuasiexperimentales, enfoques cualitativos, modelos multiniveles, etcétera.
Doctorado, maestría y especialización en medición y evaluación	Centre for the Advancement of Measurement, Evaluation, Research & Assessment, Institute for Studies in Education de la Universidad de Toronto	Formación dedicada a la medición y evaluación de programas y al desarrollo de métodos de investigación, recolección y validación de datos, redacción de informes, entrenamiento práctico.
Maestría en monitoreo, evaluación y sistemas de información	Instituto para el Desarrollo en Economía y Administración (IDEA Internacional) y Universidad de Laval	Capacitación para monitorear la realización física y financiera de los programas y proyectos; organizar y gestionar evaluaciones; y asegurar el desarrollo e implementación de sistemas de información integrales.
Posgrado en política pública y evaluación de programas	Universidad de Carleton, Ottawa	Formación en la conducción y respuesta a las evaluaciones públicas para el gobierno, los grupos de interés y las organizaciones, así como en la justificación de sus necesidades y valores.
Diploma de grado en evaluación	Facultad de Educación y Escuela de Psicología de la Facultad de Ciencias Sociales Universidad de Ottawa	Programa interdisciplinario, impartido en inglés y francés, que ofrece capacitación avanzada en evaluación de programas para los sectores público, privado y no lucrativo.
Diploma de grado en política pública y evaluación de programas	Programa Internacional de Formación en Evaluación del Desarrollo, Universidad de Carleton e Independent Evaluation Group (IEG) del World Bank Group, junto con varios patrocinadores más, Ottawa	Acreditación mundial de evaluadores profesionales con conocimientos de métodos cuantitativos (medición de impacto, encuestas) y cualitativos (grupos focales, modelos lógicos, estudios de caso).

Fuente: ReQue, 2013 y ENAP.

Una buena parte de los programas académicos disponibles se concentra en la provincia de Quebec, donde en 2012 fue creada la Red de los Universitarios en Evaluación de Quebec (Réseau québécois des universitaires en évaluation, ReQUE) para contribuir a la formación en ese campo. Uno de los primeros trabajos de ésta consistió en realizar un inventario de los cursos que se ofrecen en su provincia. Con base en él se presenta una síntesis de programas de posgrado y grado que sólo recoge los relativos a la evaluación de políticas y programas en general (Cuadro 6). Sin embargo, existen muchos otros en el marco de programas especializados en salud, educación, urbanismo, trabajo social, etcétera, y diversas universidades también ofrecen cursos aislados: un curso en la Universidad de Concordia, otro en la Escuela Nacional de Administración Pública, tres en la Universidad de Montreal, tres en la de Quebec en Montreal, y trece en la Universidad de Laval.

En conclusión, Canadá tiene un fuerte sistema de evaluación intragubernamental no obligatorio sino basado en la cultura institucionalizada entre los funcionarios, a la que ha contribuido sustantivamente la CES y sus secciones regionales, especialmente la de Quebec. Existen publicaciones esporádicas de libros y artículos de revistas, pero la comunicación en temas de evaluación se concentra en el *Canadian Journal of Program Evaluation*. También cuenta con una amplia oferta de programas universitarios de formación en el tema.

CUADRO 6
Oferta universitaria de formación en evaluación en Quebec, Canadá

Curso/programa	Universidad/escuelas	Descripción
Maestría en asuntos públicos. Análisis y evaluación	De Laval	Formación interdisciplinaria que incluye a la evaluación, entre otros contenidos.
Maestría en administración pública	Escuela Nacional de Administración Pública	Concentración en evaluación de programas.
Maestría en administración pública y política pública y Maestría en medio ambiente: curso análisis de políticas y evaluación de programas	Concordia	Visión general sobre la literatura de evaluación de programas y métodos de investigación.
Carrera de ciencias políticas: curso análisis de políticas y evaluación de programas	Concordia	Introducción a los enfoques de la planeación y evaluación de políticas públicas. Problemas conceptuales, metodológicos y éticos en la evaluación de políticas gubernamentales.
Carrera de ciencias políticas: curso métodos de evaluación de políticas	de Montreal	Métodos cuantitativos y cualitativos de análisis y evaluación, análisis comparativo, toma de decisiones, indicadores, implementación, evaluación e imputabilidad.
Diploma de estudios superiores especializados en evaluación de programas, proyectos y servicios	de Quebec en Montreal	Fundamentos del ejercicio de la evaluación exigidos por la CES para obtener la acreditación y para el análisis crítico de una evaluación. Métodos cuantitativos y cualitativos, conducta profesional, práctica de evaluación de un programa, proyecto o servicio.
Diploma de estudios superiores en administración pública	Escuela Nacional de Administración Pública	Concentración en evaluación de programas.

Fuente: ReQue, 2013 y ENAP.

América Latina

Aunque es posible encontrar algún antecedente previo, América Latina incorpora lentamente los procesos de evaluación en el último cuarto del siglo pasado, iniciando en Colombia y Costa Rica, países a los que se suma Chile y México en sus últimos años. En el nuevo siglo se han incorporado Argentina, Brasil y Perú, entre otros.

Con similar criterio al utilizado para presentar experiencias en materia de evaluación de los países avanzados de Europa y Norteamérica, se incluyen a continuación los tres países latinoamericanos que cuentan con mayor trayectoria y nivel de consolidación en sus procesos de evaluación (Colombia, Costa Rica y Chile).

CAPÍTULO V

Colombia

Myriam Cardozo Brum

Colombia no presenta programas ni publicaciones académicas, en cambio cuenta con uno de los sistemas de evaluación latinoamericanos más consolidados. Sus antecedentes en el periodo 1980-2000 se sintetizan a continuación.¹

En la década de 1980 se crea el Sistema de Evaluación y Seguimiento de Proyectos (SiSEP), enfocado a los proyectos financiados por el Banco Mundial (BM) y el Banco Interamericano de Desarrollo (BID), donde también participa el Banco de Proyectos de Inversión Nacional, encargado de las iniciativas del gobierno colombiano y dirigido a la evaluación *ex-ante*.

En 1991 se crea un sistema de evaluación de la gestión y resultados de la administración pública, encargado al Departamento Nacional de Planeación (DNP) como consecuencia de una reforma constitucional. Un año después se crea la División Especial de Evaluación y Gestión dentro del Departamento antes citado, institucionalizando las evaluaciones *ex-post*.

En 1994 se realiza la 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 y ofrecer a los ciudadanos una amplia rendición de cuentas. Pero el sistema careció del apoyo político necesario como consecuencia de la crisis de legitimidad que vivía el gobierno, en cambio, fue validado por el BM y promovido como modelo ejemplar. Al inicio

¹ Para una revisión más profunda véanse Cunill y Ospina (2003 y 2008) y Cardozo (2012a).

del nuevo siglo Sinergia había logrado ya el apoyo político suficiente para pasar a ser orgánicamente una Dirección de Evaluación de Gestión y Resultados.

Institucionalización de la evaluación

Sinergia se organiza en tres subsistemas: 1) el Sistema de Seguimiento a Metas de Gobierno (Sismeg), 2) el Sistema Nacional de Evaluaciones (Sisdeval), y 3) el Sistema de Difusión y Rendición de Cuentas, que también coordina la Red de Monitoreo y Evaluación de Política Pública (Redcolme). La estructura y órganos encargados de liderar el proceso están establecidos en la Constitución Política. Sus actividades fueron reglamentadas por el DNP mediante el Decreto 1290 de 2014.

El seguimiento de metas del Plan Nacional de Desarrollo (PND) se establece en función del presupuesto disponible. Se determinan las líneas de base y los indicadores de avances, de manera que las entidades registren en internet sus avances, restricciones, alertas y dificultades. Con esta información el DNP, en colaboración con la Presidencia de la República, elabora sus informes.

Las principales publicaciones del sistema incluyen: 1) el Balance de resultados del PND, difundido anualmente desde 1994, 2) los Informes efectuados por el presidente al inicio de las sesiones del Congreso (julio de cada año) respecto de los avances del PND, 3) el Boletín Bimestral, que desde 2010 informa sintéticamente sobre el mismo tópico a la población, y 4) el Tablero de control, que informa a los funcionarios sobre el progreso de sus metas.

En cuanto a las llamadas “evaluaciones efectivas” o “evaluaciones focalizadas”, se diseñan, ejecutan, difunden y aplican los resultados de las evaluaciones con el objetivo de minimizar costos y maximizar su utilización en la mejora continua de las políticas y programas. Para ello, Sinergia selecciona y prioriza los programas a evaluar con los diferentes sectores y en función de los recursos involucrados, el tamaño y características de la población afectada, la importancia política del sector, el nivel de innovación de la política, su potencial de aplicación y la replicabilidad de las lecciones derivadas de su evaluación. A partir del 2008 las políticas y los programas cuentan con una estructura de marco lógico que facilita el desarrollo del proceso de evaluación y permite identificar los objetivos y las estrategias del PND con que están relacionados.

Para cumplir con el proceso de evaluaciones estratégicas externas se definió un programa de evaluaciones que se extendió de 2004 a 2008. El DNP decidió iniciar con la evaluación de los programas que conforman la Red de Apoyo Social: Empleos en acción, Familias en acción y Jóvenes en acción. Esto dio lugar a realizar ocho trabajos de evaluación, en los que se trató de aplicar una estrategia experimental pero, debido a diversas modificaciones durante la implementación de los programas, se optó por un enfoque cuasiexperimental. Sus resultados fueron publicados mediante ocho números de la serie “Evaluación de políticas públicas” (Cuadro 7).

Roth (2009) resalta del Cuadro 7 que la mayoría de las evaluaciones fueron contratadas a la misma empresa, conformada por un instituto de investigación inglés dedicado al análisis de políticas sociales y económicas mediante técnicas microeconómicas y dos empresas colombianas de consultoría privada, una especializada en estudios económicos y otra en la recolección y tratamiento de datos. También que prácticamente siempre se usó la misma metodología, identificada con la perspectiva científica tradicional de fondo positivista o de “tercera generación” según Guba y Lincoln (1989). Se trató de una visión tecnocrática, de “arriba hacia abajo”, y sin participación de los beneficiarios ni de los funcionarios ejecutores.

Otro aspecto importante, señalado por Roth, es que las dos evaluaciones de Empleos en acción, que reconocían resultados positivos al programa (en comparación con no aplicarlo), fueron publicadas cuando éste ya había sido cancelado en marzo de 2003 y todos sus proyectos terminados al siguiente año. Menciona que posteriormente, los evaluadores trataron de justificar la decisión tomada por el gobierno aduciendo diferentes razones, como la existencia de alternativas, cuyos posibles impactos no fueron evaluados para justificar el cambio.

Algo similar relata Roth que sucedió con la evaluación de Familias en acción que reconoció sólo pequeños impactos en las cabeceras de los municipios, mientras que se decidió la expansión del programa; lo que mostraría que las evaluaciones se realizaron en un escenario de decisiones políticas ya tomadas, en que sus resultados no fueron considerados como elementos necesarios para asumir los cambios. El autor concluye que la práctica evaluativa oficial de Colombia nace de la alianza entre el saber científico exclusivo de los expertos y del poder político, autoritario y vertical.

CUADRO 7
Colombia: evaluaciones contratadas por el DNP

Fecha de realización	Programa o fondo evaluado	Tipo de evaluación	Equipo evaluador	Método
Marzo 2004	Familias en acción	Condiciones iniciales de los beneficiarios e impactos preliminares.	Institute of Fiscal Studies, Econometría y Sistemas Especializados de Información, SA.	Cuasi experimental
Octubre 2004	Empleos en acción	Condiciones iniciales de los beneficiarios e impactos a corto plazo.	Institute of Fiscal Studies, Econometría y Sistemas Especializados de Información, SA.	Cuasi experimental
Noviembre 2004	Red de apoyo social	Conceptualización y evaluación de impacto.	Grupo de evaluaciones focalizadas de la Dirección General de Evaluación de Políticas Públicas del DNP.	Cuasi experimental
Mayo 2006	Familias en acción	Impacto del programa a un año y medio de su ejecución.	Institute of Fiscal Studies, Econometría y Sistemas Especializados de Información, SA.	Cuasi experimental
2007	Vivienda de interés urbano	Impactos en la calidad de vida y evaluación del proceso de focalización.	Centro de Estudios Regionales Cafeteros y Empresariales.	Cuasi experimental
Enero 2008	Empleos en acción	Impacto del programa sobre los beneficiarios y sus familias.	Institute of Fiscal Studies, Econometría y Sistemas Especializados de Información, SA.	Cuasi experimental
Febrero 2008	Familias en acción	Impactos en capital humano y evaluación beneficio-costos del programa.	Institute of Fiscal Studies, Econometría y Sistemas Especializados de Información, SA.	Cuasi experimental
Octubre 2008	Fondo Colombiano de Modernización y Desarrollo Tecnológico de las Micro, Pequeñas y Medianas Empresas	Evaluación de impacto.	Econometría, SA.	Aplicación de encuestas cuantitativas a muestra de operadores del programa en 2001-2004

Fuente: reorganización del contenido de Roth (2009).

Actualmente Sinergia identifica las necesidades de evaluación, define los términos en que ésta se realizará (su tipo y metodología), contrata un evaluador externo adecuado, supervisa su trabajo, y se compromete con la socialización y el uso efectivo de los resultados mediante la definición de un plan de mejoras. Los temas en que se producen más contrataciones incluyen el diseño de sistemas de monitoreo y evaluación, la preparación de TR, el diseño de indicadores, la evaluación de impacto y de políticas públicas, entre otros. Los ministerios no prestan estos servicios porque cuentan con limitadas capacidades internas, lo hacen las universidades (Javeriana y de los Andes), fundaciones, consultoras y organismos internacionales ya que han desarrollado suficiente personal capacitado en estos temas. Esto vuelve relativamente fáciles las contrataciones, salvo en temas muy especializados en que su costo puede ser muy elevado para los presupuestos públicos disponibles.² La ley establece la obligación de que las contrataciones se hagan mediante licitación pública. El proceso se desarrolla bajo un control técnico estricto de cada una de estas fases. También pueden contratar los propios ministerios y el DNP (Quintero, 2010).

Los profesionales calificados suelen ser economistas con maestrías, pero mientras hay muchos econométricos, falta capacidad de análisis político y manejo metodológico. Se presentan más problemas en la comunicación de resultados que en la recolección y análisis de información (Quintero, 2010).

Hasta el 2012, Sinergia reporta haber realizado 171 evaluaciones de políticas, programas, sistemas y procesos en los más diversos sectores (salvo salud) como: educación, vivienda, turismo, transporte, infraestructura, desarrollo rural, trabajo, medioambiente, agua, violencia de género, seguridad alimentaria, discapacidad y, en general, lucha contra la pobreza.³ En los pocos casos en que se especifica el tipo de evaluación se incluyen las cualitativas, de procesos, de resultados, de impacto, ejecutivas, de mediano término y finales; también

² Una evaluación de impacto puede costar entre 350 000 y 500 000 dólares.

³ En 2012 Sinergia informa en su portal electrónico la realización de seis evaluaciones: 1) de impacto de Familias en acción; 2) de resultados de Infraestructura de Acción social, 3) institucional y de operaciones de Producción de la Información Hidrometeorológica y Ambiental, 4) de resultados de Apoyo al Sector Agua Potable y Saneamiento Básico, 5) de operaciones y resultados de Corredores Arteriales de Competitividad, y 6) de operaciones de Sistemas de Vigilancia Electrónica.

se menciona el levantamiento de líneas de base y el desarrollo y cálculo de indicadores (especialmente indicadores cualitativos de desempeño).

En 2013, Sinergia reporta en su página 16 evaluaciones acabadas de los diversos tipos, donde sólo dos incluyen impacto (Programa de Consolidación de Vivienda de Interés Social Urbana y Concesiones de Aeropuertos, Puertos y Líneas Férreas). En 2014 sólo aparecen cuatro terminadas, incluyendo una de impacto (Fondos de Desarrollo y Compensación Regional) pero se reportan 17 más en curso. En los dos años mencionados casi la mitad de los trabajos han sido realizados por una misma consultora.⁴

La Ley 617 del 2000 también responsabiliza al DNP de la evaluación territorial a nivel de los municipios, para lo que éste ha definido una metodología integral y un conjunto de indicadores que miden la eficacia, la eficiencia y la efectividad del desempeño, principalmente en salud, educación, alimentación y servicio de agua potable. Los resultados se publican anualmente por internet y mediante la distribución de cuatro mil boletines que se distribuyen entre los funcionarios y la población.

Quintero (2010) concluye que a pesar de la larga trayectoria de Sinergia y de la calidad de los profesionales que ofrecen sus servicios evaluativos, el país tiene un déficit importante en materia de evaluación de impactos, enfoque de derechos y análisis sistémico, entre otros. Agrega que se requiere superar las restricciones estructurales de corte político, institucional, económico y cultural que impiden considerar al monitoreo y a la evaluación como elementos fundamentales de una gestión pública que aporta a la transparencia y la rendición de cuentas como marca la Constitución colombiana.

Indudablemente el país ha avanzado en consolidar su sistema nacional de evaluación con la generación de una cultura orientada a resultados y rendición de cuentas (aunque falta aún mucho a nivel de gobernaciones y alcaldías), un desarrollo importante de los instrumentos y una definición clara de las responsabilidades. La evaluación ha cambiado su orientación: de la eficiencia del gasto a los resultados e impactos. En cambio, entre sus debilidades también se encuentran: la poca utilización de las conclusiones de las evaluaciones en los

⁴ Unión Temporal Econometría Consultores-SEI, SA.

procesos decisorios, su limitada credibilidad a nivel social y la casi ausencia de trabajos de metaevaluación.⁵

Profesionalización de la evaluación

Asociaciones profesionales

En 2007 Colombia fue sede de la II Conferencia de la Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (Relac), pero carece de una asociación de profesionales independientes. En cambio, la Red Colombiana de Monitoreo y Evaluación (Redcolme) cuenta con 1 300 miembros y tiene como objetivo fortalecer la cultura de gestión por resultados, especialmente a nivel local de gobierno, mediante la generación de espacios físicos y virtuales de cooperación horizontal y vertical entre gobiernos territoriales, entidades nacionales, academia y centros de pensamiento. En noviembre de 2013 realizó su IV Encuentro, que significó el relanzamiento de la red, con el objetivo de incentivar el diálogo de saberes y la cualificación de los actores e interesados en los temas de seguimiento y evaluación de políticas públicas. Forma parte de la Red Latinoamericana y del Caribe de Monitoreo y Evaluación (Redlacme).

Publicaciones

Casi no se identificaron publicaciones en revistas especializadas y libros. Sólo se ha localizado una revista que esporádicamente publica algún artículo sobre el tema. Es el caso de “La evaluación y el análisis de políticas públicas”, de Carlos Salazar Vargas, incluido en el número 9 de la revista *Opera*, Universidad Externado de Colombia, 2009. El mismo autor publicó en 2012 el libro *De la evaluación al análisis de las políticas públicas*, en la Editorial Académica Española. Otro trabajo a nivel local y del sector salud lo constituye *Evaluación de la aplicación*

⁵ Sólo se localizó una tesis de un *magister* en educación, Jorge Mahecha (2007), y un libro de Álvarez y Niño (2007).

de políticas públicas de salud en la ciudad de Cartagena, publicado en 2006 por Eumednet en Cartagena de Indias, Colombia.

Programas de formación

Si bien la mayoría de las instancias gubernamentales le otorga importancia a la capacitación, la oferta de diplomados y cursos de especialización es limitada, por lo que sólo se permite la asistencia de dos o tres personas por ministerio a cada curso. Por su parte, la oferta académica no incluye posgrados especializados en evaluación sino que se concentra en los de economía, ingeniería, sistemas de gestión y control organizacional, relaciones internacionales y ciencias políticas, en las Universidades de los Andes, Nacional y Externado de Colombia principalmente, los que se consideran insuficientes y atendidos por profesores poco actualizados (Quintero, 2010).

Por ejemplo, se menciona la primera Jornada de Actualización en Evaluación organizada el 13 y 14 de noviembre pasado por el Grupo de Investigación en Evaluación de Planes, Programas y Proyectos en el Campo de la Alimentación y Nutrición Humana (Gepanh), el Programa de Especialización en Alimentación y Nutrición en Promoción de la Salud y el Proyecto de Extensión en Interventoría a Servicios de Suministro de Alimentación; adscritos al Departamento de Nutrición Humana de la Universidad Nacional de Colombia. En el mismo se planeó la participación de expertos internacionales, como un investigador del Instituto Nacional de Salud Pública de México (INSP), quien expondría el tema de Evaluación de impacto e impartiría el seminario Metodologías de evaluación de impacto de programas de salud y nutrición para exponer un panorama general de los diseños de evaluación de impacto experimentales y no (cuasi) experimentales y las técnicas más utilizadas.

A modo de conclusión sobre el caso colombiano es necesario reconocer que ha sido pionero en América Latina y cuenta con un sistema consolidado de evaluaciones externas, que no muestra evidencias claras sobre la utilización de sus recomendaciones y no siempre ha garantizado plenamente la independencia de los órganos contratados para realizarlas. García y García (2010) coinciden en ubicarlo con un alto grado de institucionalización de su sistema y un nivel

medio de consideración de sus resultados en la toma de decisiones (lo califican con 3.7 en una escala de 0 a 5, frente al 1.6 del promedio de América Latina y el Caribe). Tiene serios problemas para formar y capacitar especialistas por falta de programas universitarios suficientes, especialmente en evaluación de impacto y metaevaluación, y ausencia de espacios asociativos de reflexión y discusión entre colegas.

CAPÍTULO VI

Costa Rica

Myriam Cardozo Brum

Costa Rica es uno de los países latinoamericanos pioneros en materia de evaluación. Los principales eventos ocurridos entre 1974 y 2000 se presentan a continuación.¹

En 1974 se promulgó la Ley de Planificación Nacional y en 1982 la Oficina de Planificación Nacional se transformó en el Ministerio de Planificación Nacional y Política Económica (Mideplan), encargado de la evaluación sistemática de los resultados de los planes, políticas y programas de desarrollo.

Veinte años después el Plan Nacional de Desarrollo 1994-1998 (PND) creó un sistema de evaluación de la gestión y un área específica de Evaluación y Seguimiento, que se acompañó de un decreto del Ejecutivo que da origen al Sistema Nacional de Evaluación (Sine), cuyos alcances fueron redefinidos en 1995. Este sistema, en parte inspirado en la experiencia colombiana, surgió con un fuerte respaldo de la Presidencia de la República para alcanzar una gestión basada en resultados y se ha mantenido a pesar de los cambios de administración.

En el año 2000 la nueva Constitución incorporó la evaluación por resultados y la rendición de cuentas como principios inspiradores de la institucionalidad democrática, los que ya presentaban un antecedente importante en la Constitución de 1949.

¹ Para una revisión más profunda véanse Cunill y Ospina (2003 y 2008) y Cardozo (2012a).

Institucionalización de la evaluación

Mideplan encabeza el esfuerzo burocrático con un apoyo importante de la cooperación alemana, especialmente la Universidad del Sarre. Entre sus productos figura la definición anual de lineamientos y metodologías de seguimiento y evaluación, así como la presentación de informes. En estos documentos se encuentran desde definiciones relevantes hasta la construcción de indicadores relativos a gestión y desempeño con costos, resultados, población usuaria y su distribución regional. Mideplan cuenta con una Unidad de Acciones Institucionales que realiza estas funciones en relación con el PND y otra de Acciones Estratégicas que lo hace en torno a las políticas y programas gubernamentales seleccionados. También asesora y capacita al personal del Sine en el mejoramiento de la gestión pública.

El Sine tiene dos módulos: el de autoevaluación (o automonitoreo) y el de evaluación estratégica. Originalmente se desarrolló el primero; su principal instrumento, aplicado en forma semestral y anual, es la Matriz Anual de Programación y Seguimiento Sectorial e Institucional, creada en 2010. Cuando en ésta se detectan incumplimientos importantes se elabora un Plan de Mejoras. En 2013 se trató también de lograr la verificación del 5% de la información ofrecida por los diferentes sectores dentro de este proceso.

El “Informe Anual de la Matriz de Programación y Seguimiento Sectorial e Institucional” de 2013 menciona el mejoramiento de los servicios públicos logrado con la contribución del área de Evaluación y Seguimiento. También la aplicación de una encuesta de percepción de las unidades de planificación institucional, con el apoyo de la Universidad de Costa Rica, cuyo informe estaba previsto para el 2014.

En cuanto a las evaluaciones estratégicas, en 2006 se definieron sus métodos, tipos, plazos y destinatarios y en 2012 su enfoque dirigido a evaluar políticas, planes, programas y proyectos quedó establecido en el “Manual gerencial para el diseño y ejecución de evaluaciones estratégicas de gobierno”. Las iniciativas de evaluación provienen de las propias instituciones y se seleccionan en función de los recursos disponibles. En 2011-2014 se habían programado cuatro evaluaciones (una en cada año) que enfrentaron algunos problemas en su implementación, pero todas culminaron en 2014. Además, se informa un avance satisfactorio en tres nuevas evaluaciones de efectos.

Las evaluaciones concluidas fueron:

1. Evaluación del diseño metodológico y proceso de elaboración del Plan Nacional de Desarrollo 2011-2014 (Mideplan). Entre sus conclusiones incluyó: reorientar el diseño hacia resultados más estratégicos y evaluables, aumentar su vinculación con el presupuesto y facilitar la participación-representación más diversa de actores sociales e institucionales.
2. Evaluación del Proyecto de Desarrollo Organizacional, Empresarial, Asociativo y Comunitario La Germinadora, que deberá seguir impulsando la capacitación, la promoción y el seguimiento para generar ingresos mediante iniciativas empresariales en zonas en condiciones de pobreza (Instituto Mixto de Ayuda Social, IMAS).
3. Evaluación del Proyecto Apoyo a la Gestión Pedagógica de Centros Educativos con Orientación Inclusiva, que reconoce el esfuerzo realizado al pasar de seis a 10 de estos centros entre 2009 y 2011 (Ministerio de Educación Pública, MEP).
4. Evaluación del Distrito de Riego Arenal-Tempisque, Secretaría de Planificación del Sector Agropecuario (Sepsa) y Servicio Nacional de Aguas Subterráneas, Riego y Avenamiento (Senara).

Pero el énfasis costarricense está en la autoevaluación realizada por las propias instituciones gubernamentales, para cuya realización el Sine cuenta con una guía detallada que gira en torno de los Compromisos de resultados que aquellas han asumido a partir de un proceso de negociación. Con base en su aplicación, el sistema elabora un informe anual, conjuntamente con el Ministerio de Hacienda, que incluye las calificaciones obtenidas por las instituciones y sectores en el cumplimiento de sus compromisos, un análisis de los factores de éxito y las dificultades presentadas. En 2001, la ley estableció que estos informes deberían avanzar gradualmente tanto para lograr una cobertura nacional, regional, sectorial e institucional, como para obtener los distintos tipos de evaluaciones previstas: de cumplimiento de metas, eficacia, eficiencia, impactos y medición de costos.

Durante los últimos años, el Sine ha consolidado una red de información de instituciones del sector público y ha logrado establecer mecanismos de coordinación entre Mideplan, el Ministerio de Hacienda y la Contraloría

General de la República (2011) para vincular a la evaluación con los procesos de planificación, programación y ejecución presupuestaria.

El Sine dispone también de un sistema de incentivos para que los resultados de las evaluaciones sean considerados, de manera de promover el ahorro público y alcanzar los niveles de eficiencia social deseados, costo, calidad, y cobertura. Pretende que dichos resultados constituyan la base para otorgar el presupuesto a las instituciones públicas y los incentivos a los funcionarios responsables de la ejecución de los programas, proyectos y acciones; sin embargo, ha mantenido demasiado desvinculado el sistema de evaluación y el de asignación presupuestal. Por otro lado, a pesar de que la normatividad no exige la utilización instrumental, es importante subrayarla, ya que de las 30 recomendaciones dirigidas a la preparación del PND, 27 fueron recuperadas en la propuesta metodológica para el PND 2015-2019. Las tres restantes no fueron incluidas porque Mideplan las consideró fuera de su competencia o que ya estaban contempladas en otros de sus procesos de trabajo.

Díaz (2005), al metaevaluar al Sine, reconoce sus aportes pero también identifica un conjunto de debilidades: importancia desproporcionada atribuida al PND como instrumento de evaluación, excesiva simplificación mediante la medición numérica de cumplimiento de acciones estratégicas, instrumento evaluativo insuficiente frente al tamaño y complejidad de lo evaluado, rigidez del instrumento en relación con el tiempo (máximo los cuatro años de gobierno), notable ausencia de mecanismos de verificación y validación, descuido de la participación de los funcionarios en el proceso, insuficiente sistema de tratamiento automatizado de información, entre otros.

Mideplan también ha contratado consultorías como la “Revisión experta al personal de Mideplan en metodologías para seguimiento y evaluación del PND y conformación del Registro Nacional de Evaluaciones Estratégicas del Gobierno (2010-2011)” con financiamiento de la Agencia Española de Cooperación Internacional para el Desarrollo (AECID) y el Programa de Naciones Unidas para el Desarrollo (PNUD), que analiza la posibilidad de adoptar un sistema similar al de la Ciudad de Navarra en España.

En síntesis, Costa Rica enfrenta aún resistencias importantes a la evaluación por temor a posibles sanciones y por limitaciones como las descritas por Díaz (2005), por lo que falta institucionalizar plenamente una cultura que permita a la sociedad valorar su utilidad como herramienta de gestión y aumentar su

credibilidad, así como convencer al gobierno para realizar un mayor número de ellas en forma externa y con metodologías más amplias. Sin embargo, es necesario reconocer la utilización de un dispositivo original (incentivos) para lograr un avance gubernamental importante en materia de utilización de las recomendaciones recibidas.

Profesionalización de la evaluación

Asociaciones profesionales

La Asociación Centroamericana de Evaluación, integrada fundamentalmente con especialistas de Costa Rica, impulsa la evaluación de programas y proyectos sociales y privados. En 2010 Costa Rica fue sede de la III Conferencia de la Relac y en 2012 de la VIII Conferencia Anual de la Redlacme, esta última organizada por el Banco Mundial, el Banco Interamericano de Desarrollo y el Ministerio de Planificación Nacional y Política Económica.

Publicaciones

Costa Rica no cuenta con revistas especializadas en la evaluación de políticas y programas públicos, sólo una revista electrónica de *Actualidades investigativas en educación*, que regularmente publica artículos sobre el tema. La Asociación Centroamericana de Evaluación publica también un Boletín Bimensual Electrónico que se distribuye entre 1 400 suscriptores latinoamericanos. Otros boletines electrónicos son difundidos por el Sine y el Programa de Fortalecimiento de las Capacidades en Evaluación (Foceval). En temas cercanos a la evaluación puede mencionarse la *Revista CCAalidad*, del Consejo Centroamericano de Acreditación de la Evaluación Superior.

En 2009 se tradujo y publicó un libro que había sido publicado en Alemania cuatro años antes (Stockmann, 2009), dos años después la Universidad de Costa Rica publicó un manual del mismo autor (Stockmann, 2011) y más recientemente se ha difundido otro de Mideplan (2012).

Programas de formación

En el caso de Costa Rica, la formación en evaluación es impulsada por el propio gobierno, además de las actividades realizadas por las instituciones académicas.

Desde 2005 Mideplan ha ofrecido cursos de evaluación de capacitación profesional, los que se consolidaron en 2011 cuando inició el Foceval para Costa Rica y Centroamérica, con el apoyo de la Agencia Alemana para el Desarrollo. Su objetivo es capacitar a oferentes y demandantes de servicios de evaluación para que mejoren sus capacidades de análisis de información y de exposición de conclusiones que contribuyan a una mejor adopción de decisiones públicas y a la rendición de cuentas. Su planeación se llevó a cabo en forma participativa y se establecieron metas hasta el 2015. Entre sus productos se encuentra un curso básico de capacitación de enlaces sectoriales e institucionales en diseño y ejecución de evaluaciones estratégicas y un taller de inducción sobre informes de seguimiento, desarrollados en 2013.

Para promover la cultura de aprendizaje y la mejora institucional, Foceval ha promovido la realización de evaluaciones estratégicas piloto a partir de las iniciativas propias de entidades de gobierno puestas a concursar y la realización de licitaciones. En abril de 2014 se habían concluido las cuatro ya mencionadas, desarrolladas con la metodología de “aprender haciendo”.

En forma similar a lo sucedido con la utilización de las recomendaciones en el PND, en el marco del MEP, tres recomendaciones se asumieron plenamente, dos parcialmente, dos se postergaron, dos se rechazaron por no encontrarse dentro de sus atribuciones. Mientras que en el IMAS se han aplicado algunas, pero aún se está en proceso de valoración de la mayoría para su posible implementación.

Foceval-Mideplan también ha contratado consultorías como la financiada por la Agencia Alemana de Cooperación Internacional (Gesellschaft für Internationale Zusammenarbeit, GIZ) en 2011-2012 para diseñar instrumentos metodológicos para sistematizar y fortalecer los procesos de evaluación estratégica de políticas, planes, programas y proyectos.

Por otra parte, Costa Rica y Chile son los únicos países latinoamericanos que ofrecen una maestría en evaluación. En el caso costarricense se trata de la Maestría en Evaluación de Programas y Proyectos de Desarrollo impartida por las Escuelas de Trabajo Social y de Administración Pública; también

existen cursos en otros programas como el Master Profesional en Población y Salud, ambos en la Universidad de Costa Rica. Si bien la primera es de carácter profesionalizante, en sus ya 20 años de existencia cuenta con bastante producción académica, lo que ha facilitado la publicación de algunos libros y artículos, la presentación de ponencias y la realización de tesis (Acuña, Alarcón, Solano; Vargas, Morillo, Calvo, etcétera),² incluso en temas de metaevaluación (Díaz, 2002; Centeno, 2002; Díaz y Rosales, 2003; Díaz, 2005). El programa se desarrolló originalmente durante un periodo exploratorio, tanto en estructura como en contenido, con el apoyo de la cooperación canadiense; posteriormente logró su consolidación y proyección internacional con la participación de teóricos de la talla de House o de Scriven; y actualmente se ha acercado al enfoque sociológico que promueve la cooperación alemana del Sarre y a la visión del Ministerio de Planificación Nacional y Política Económica.

A nivel académico también existen esfuerzos poderosos en materia de evaluación educativa y movimientos de acreditación de la educación superior que promueven una visión orientada a la rendición de cuentas. Finalmente, se considera que Costa Rica ha sido pionero en América Latina en su esfuerzo de evaluación interna y que su nivel de utilización de las recomendaciones resulta muy meritorio, pero su avance ha seguido un proceso lento, que muestra un reducido número de evaluaciones anuales y diversos problemas aún no resueltos. García y García (2010) lo ubican como un país de avance medio en materia de insitucionalización de su sistema y conceden un menor reconocimiento a la utilización de sus resultados para la toma de decisiones, ubicándolo también a nivel medio (lo califican globalmente con 3.0 en una escala de 0 a 5, frente al 1.6 del promedio de América Latina y el Caribe). También es de reconocerse que cuente con un posgrado en evaluación y una asociación donde discutir los problemas vinculados a su teoría y práctica, sólo sería conveniente avanzar en la disposición de publicaciones periódicas sobre el tema.

² Citados en “Lista de trabajos finales de graduación”, vol. 1, sin especificar fecha [sibdi. ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).

CAPÍTULO VII

Chile

Myriam Cardozo Brum

Desde comienzos de la década de 1990 los gobiernos chilenos han emprendido serios esfuerzos por desarrollar un mecanismo de evaluación gubernamental. A continuación se presenta una síntesis de los hechos más significativos que contribuyeron a su institucionalización durante esa década.

En 1990 la Oficina de Planificación Nacional se transformó en el Ministerio de Planificación y Cooperación, encargado de evaluar los proyectos de inversión financiados por el Estado, que recomendó la utilización voluntaria de indicadores de desempeño junto con el Sistema de Metas Ministeriales que se mantuvo vigente hasta 1999. En 1994 el gobierno creó un Sistema de Control de Gestión y Presupuestos por Resultados.

En 1996 se preparó el Plan Estratégico de Modernización de la Gestión Pública que contiene los instrumentos de evaluación de los programas gubernamentales, cuya aplicación inició al siguiente año mediante la creación del Sistema de Evaluación de Programas Gubernamentales y la implantación de los Balances de Gestión Integral. Se trató del primer intento por articular sus instrumentos de evaluación (de metas ministeriales, desempeño de funcionarios, etcétera) en un sistema integrado. Los indicadores de desempeño se volvieron obligatorios, pero su uso fue interrumpido en 1999.

En 1998 se incorporó el mecanismo de incentivos y remuneraciones institucionales, que busca fomentar la adopción de las recomendaciones de las evaluaciones y darle seguimiento a su cumplimiento.

En el año 2000 se aprobó el Proyecto de “Reforma y Modernización del Estado” cuya segunda línea de acción se refiere a la evaluación de los progra-

mas públicos, que a partir de ese momento tiende a integrarse en un solo sistema: Sistema de Control de Gestión (SCG). Los indicadores de desempeño, metodológicamente ajustados, se volvieron nuevamente obligatorios.

Institucionalización de la evaluación

En un principio la práctica evaluativa chilena tuvo su fundamento en la preocupación por el uso adecuado de los recursos públicos y por la alimentación del proceso presupuestario, por lo que se componía de instrumentos orientados al control que, como ya se mencionó, no formaban parte de un sistema único.

En 2001 y 2002 se introdujeron nuevos tipos de evaluación: los de impacto de programas y la evaluación comprehensiva del gasto por instituciones.¹ Ese último año se firmó el acuerdo entre el Poder Ejecutivo y el Congreso que permitió el nacimiento del Sistema de Evaluación y Control Gubernamental (SECG) destinado a mejorar la asignación y el uso eficiente de los recursos públicos a cargo de la Dirección de Presupuestos (Dipres) del Ministerio de Hacienda, en coordinación con el Mideplan. La práctica evaluativa se dotó de carácter sistemático, metodológico e instrumental, y se establecieron los criterios de forma y fondo de la información producida. El gobierno diseñó mecanismos tanto para promover las evaluaciones como para utilizar su contribución en la toma de decisiones públicas. En particular, el gobierno incorporó tres nuevas líneas de acción que han permitido situar a este sistema en la vanguardia (Cunill y Ospina, 2008):

- Aumento en la cobertura de las evaluaciones de programas e instituciones a partir del 2006, lo que se traduce en uno de cada dos programas presupuestarios, sin importar su rubro.
- Entrega de un paquete de servicios de asistencia técnica y metodológica² a la evaluación a partir de 2007.

¹ Para mayor información véanse Cardozo (2012a) y Hernández (en proceso).

² Esta asistencia técnica y metodológica es brindada por medio de cursos de capacitación y actualización de metodologías evaluativas que están dirigidas a las áreas responsables de coordinar las evaluaciones en cada uno de los ministerios.

- Creación en 2009 de una nueva línea de evaluación, denominada “Evaluación de Programas Nuevos (EPN)”.

Las evaluaciones son realizadas por consultores externos organizados en paneles de expertos, universidades o consultoras privadas, que son seleccionados en concursos públicos. Las instancias evaluadas proveen información, participan en reuniones y comentan los informes, pero no constituyen la contraparte, ni técnica ni administrativa. La información de los indicadores de desempeño por institución es enviada por la Dipres al Congreso y éste la utiliza para elaborar el presupuesto del siguiente año y asignar los incentivos que el sistema brinda tanto a los ministerios como a los funcionarios (Guzmán, 2009). El Cuadro 8 muestra los procesos relativos a la evaluación de programas que el sistema incluye.

Las evaluaciones de impacto se refieren a procesos y resultados (en sentido amplio) y las de diseño no evalúan la formulación del programa sino que preparan el camino para realizar futuras evaluaciones de impacto. El desarrollo de las evaluaciones implica una extensa lista de etapas con la participación de actores diferentes, las que se sintetiza en el Cuadro 9.

Tal vez las fases del proceso no sean en sí originales, pero lo es el nivel de involucramiento de la Dipres con otras instancias (Congreso, otros ministerios, institución evaluada, etcétera) en la toma de decisiones dentro del proceso, desde la selección misma de los programas a evaluar.

Con el objetivo de asegurar la mayor apropiación y realización de las recomendaciones, a partir de 2008, éstas son analizadas por profesionales de las instituciones y ministerios responsables de los programas evaluados junto con personal de la Dipres. Este proceso acaba con la definición de compromisos institucionales entre esta dirección y el ministerio responsable del programa, señalando la forma particular en que se dará respuesta a los problemas detectados, su clasificación (ajustes menores, modificación en el diseño, rediseño sustantivo del programa, reubicación institucional o finalización-reemplazo íntegro del programa) y los plazos para realizarlos (Hernández, en proceso).

Todo esto se integra en un plan de trabajo, en el que se fijan responsables, etapas y plazos para cumplir las acciones establecidas, posteriormente la Dipres realiza una revisión del cumplimiento de los compromisos calificándolos como “cumplido”, “parcialmente cumplido” y “no cumplido”, de acuerdo con los medios de verificación presentados por las instituciones (Dipres, 2008)

CUADRO 8
Chile: Sistema de Evaluación y Control de Gestión

Tipo de evaluación	Descripción	Aplicaciones
Evaluación comprehensiva del gasto	Analiza la consistencia entre los objetivos organizacionales y sectoriales, el diseño, la gestión y utilización de recursos, así como la provisión de bienes y servicios de una institución o un conjunto de éstas.	Se aplica desde 2001 y en 2013 se contaba con 54 efectuadas.
Evaluación de programas gubernamentales	Utiliza el marco lógico para analizar la consistencia del diseño y los resultados de la gestión con los objetivos de los programas, con base en la información disponible. Su ejecución está a cargo de paneles de la sociedad civil compuestos por un experto en evaluación, otro en el tema a evaluar y un tercero en gestión de programas públicos. Son seleccionados mediante concurso público. Aplican métodos cualitativos sencillos y entrevistas a actores clave. Estas evaluaciones duran entre cuatro y seis meses e incluyen información sobre: justificación, diseño, organización y gestión, resultados, sostenibilidad/continuidad, así como conclusiones y recomendaciones.	Entre 1997 y 2003 se evaluaron 145 programas con esta metodología. En años posteriores se ha reducido a 15-16 en promedio por año, alcanzando en 1997-2013 un total de 299.
Evaluación en profundidad de programas públicos	Profundizan en aspectos que complementan al tipo anterior, con recolección de información primaria en campo, su procesamiento y análisis. Implican mayor volumen de información, tiempo y costos. Son realizadas por instituciones externas (universidades, consultorías).	Hasta fines de 2003 se habían evaluado 10 programas en esta modalidad.
Evaluación de programas nuevos	Implica el diseño de la evaluación desde el inicio del programa, preferentemente de manera experimental, con grupo de control, identificando la línea de base. Cuenta con el apoyo técnico de un Panel Asesor Internacional, conformado por académicos de gran prestigio y experiencia en materia de diseño e implementación de evaluaciones de impacto, y del Centro de Microdatos del Departamento de Economía de la Facultad de Economía y Negocios de la Universidad de Chile. Concluye con un informe detallado de comentarios y recomendaciones.	Comenzó con el diseño de la evaluación de cinco programas comprometidos en el Protocolo de Acuerdo que acompañó el despacho de la Ley de Presupuestos 2009. Hasta 2013 se habían realizado 14 evaluaciones de este tipo.
Evaluación de impacto de programas	Pueden implicar sólo la evaluación de resultados de corto, mediano y largo plazo, o cubrir también otros aspectos: justificación, diseño, organización y gestión, sostenibilidad/continuidad, eficiencia y economía en la gestión de recursos. Además deben presentar conclusiones y recomendaciones. Representan un volumen importante de recursos.	Se aplican a partir de 2001 a programas que ya han sido evaluados en los que no se pudieron medir los resultados finales o impacto. Incluso, se han realizado 107 evaluaciones de impacto hasta 2013.

Fuente: elaboración propia a partir de CLAD (2012), Guzmán (2009) y Dipres (2013).

CUADRO 9
Chile: diseño operativo de las evaluaciones

Etapa	Actores
Selección de programas o instituciones a evaluar	La selección se efectúa con la participación del Congreso Nacional, sobre la base de una propuesta presentada por la Dipres, elaborada con la participación del Ministerio de Planificación y la Secretaría de la Presidencia.
Selección de consultores o empresas, incluyendo universidades, mediante concurso público	Los concursos públicos los organiza la Dipres con el apoyo de comités de selección integrados también por representantes de los Ministerios de Planificación y Secretaría de la Presidencia.
Preparación de información base por las instancias a evaluar	Las unidades organizacionales son coordinadas por la Dipres.
Transmisión a los evaluadores de aspectos críticos del objeto a evaluar para que sean revisados con atención	Se efectúa una o más reuniones en que participan los evaluadores, los responsables del programa y la Dipres. También se pueden efectuar reuniones con el Congreso.
Capacitación y preparación de marcos lógicos en evaluaciones de escritorio y socialización de la metodología en las de impacto	La capacitación tanto para los evaluadores como a los responsables de los programas la efectúa la Dipres y la preparación de los marcos lógicos la realizan los evaluadores en conjunto con la unidad evaluada, bajo la coordinación de la Dipres.
Evaluación: inicio, entrega de informes de avance y final, los que son comentados por la institución objeto de evaluación y la Dipres	Es realizada por los evaluadores externos; su contraparte técnica y administrativa corresponde a la Dipres. Las unidades organizacionales evaluadas participan a través de la entrega de información, en reuniones de trabajo y revisión de informes.
Análisis de las evaluaciones terminadas	Los análisis los efectúan las unidades organizacionales evaluadas que elaboran una Respuesta Institucional. También los hace la Dipres con especialistas presupuestarios, especialistas a cargos de las evaluaciones y el director de Presupuestos para identificar las orientaciones de los cambios necesarios y sus efectos presupuestarios.
Envío oficial del informe al Congreso y presentación en sitio web	El envío del informe final al Congreso lo efectúa el director de Presupuestos, acompañado de la Respuesta Institucional y de un informe ejecutivo preparado por la Dipres, que también lo publica en el sitio web.
Envío al Congreso del Informe de Finanzas Públicas junto con el Proyecto de Presupuestos	El envío lo efectúa el Director de Presupuestos.
Formalización de compromisos sobre la base de las recomendaciones efectuadas por los evaluadores	Los compromisos se establecen entre la Dipres y la institución evaluada.
Seguimiento semestral de los compromisos institucionales	El seguimiento de compromisos de mejoramiento lo efectúa la Dipres con base en la información enviada por los organismos evaluados.
Informe de cumplimiento de compromisos y análisis de información del seguimiento	El análisis de la información del seguimiento es efectuado en la Dipres.
Envío al Congreso dos veces al año del Informe de Cumplimientos de Compromisos Institucionales	El envío al Congreso del Informes de Evaluación de la Gestión Financiera del Sector Público lo efectúa el Director de Presupuestos.

Fuente: adaptado de Guzmán (2009).

Existe también un conjunto de mecanismos para promover el uso de los resultados de las evaluaciones que se analizará en la tercera parte del libro, incluyendo el ya mencionado uso de incentivos. En cuanto a la difusión de las evaluaciones, el sistema cuenta con el informe llamado “Balance de la Gestión Integral (BGI)”. Es el instrumento mediante el cual cada institución da a conocer con datos extraídos de las evaluaciones los resultados de su gestión anual al Congreso Nacional, a la Dipres, los otros ministerios y al público en general (Hernández, en proceso).

Profesionalización de la evaluación³

Asociaciones profesionales

Si bien no es específica en evaluación, existe la Asociación Chilena de Políticas Públicas, integrada por profesionales con interés y práctica en los campos del seguimiento, evaluación y sistematización de políticas, programas y proyectos de desarrollo. La Relac tiene bastante presencia en el país y algunos evaluadores integran también la Red Latinoamericana y del Caribe de Monitoreo y Evaluación.

Publicaciones

Aunque no se refiere a los procesos de evaluación en general ni es exclusivamente chilena, existe una amplia participación nacional en la *Revista Iberoamericana de Evaluación Educativa*, editada periódicamente por la Red Iberoamericana de Investigación sobre Cambio y Eficacia Escolar (Rinace). También existen otras que, aunque no son periódicas ni específicas, incluyen artículos relacionados con el tema en más de la mitad de sus publicaciones: 1) *Revista INVI* del Instituto de la Vivienda de la Universidad de Chile, que elabora evaluaciones de políticas y programas de vivienda, servicios públicos y pobreza urbana, 2) *Ciencias Política y Administración Pública*, que incluye entre sus trabajos sobre política y gestión pública algunos relativos a la cultura de la evaluación desde

³ Preparado a partir de la información de Hernández (tesis en proceso).

un enfoque teórico, 3) *Universum. Revista de Humanidades y Ciencias Sociales* de la Universidad de Talca, que recoge el estudio de metodologías evaluativas, y 4) publicaciones semestrales del Núcleo de Evaluación de Políticas Públicas, perteneciente a la Facultad de Ciencias Sociales de la Universidad de Chile, entre otras. En este momento la *Revista Análisis Público* de la Universidad de Valparaíso está convocando a enviar artículos a publicarse en 2015. Todas las revistas mencionadas pueden accederse en sus sitios de internet.

No se identificaron libros publicados sobre evaluación en general, sólo trabajos aplicados en educación y medio ambiente o bien de políticas públicas con un capítulo sobre el tema.

Programas de formación

Chile ofrece una serie de maestrías y diplomados en evaluación dentro de diversas universidades y centros de investigación, algunas específicas en temas educativos o ambientales. A continuación se enlistan las principales (Cuadro 10).

CUADRO 10
Chile: principales programas de formación en evaluación

Universidad/otras organizaciones	Programa
Universidad de Santiago de Chile	Maestría en evaluación de la gestión pública
Pontificia Universidad Católica de Chile	Maestría en medición y evaluación de programas educativos
Universidad Finis Terrae	Maestría en medio ambiente: gestión, evaluación y derecho ambiental
CEPAL, Chile	Diplomado en evaluación de la gestión y programas públicos
Universidad de Chile	Diplomado en evaluación de impacto de políticas públicas
Universidad de Chile	Diplomado en diseño, evaluación y gestión de proyectos de interés público.
Universidad Andrés Bello	Diplomado en metodologías para evaluar intervenciones sociales
Flacso Chile	Diplomado en evaluación y monitoreo de políticas y programas sociales
Universidad de Lagos	Diplomado en evaluación de impacto de políticas ambientales
Universidad Alberto Hurtado	Diplomado en formulación y evaluación de programas y proyectos. Metodología del marco lógico

Fuente: adaptado de Hernández (en proceso) con datos del Ministerio de Educación de Chile.

A pesar de su incorporación un poco tardía y de las dificultades encontradas en el proceso, Chile ha logrado avanzar en la institucionalización de su sistema de evaluación, transitando de una orientación de eficiencia en el gasto a una con énfasis en los resultados e impactos. Implicó la construcción de un sistema integral que ha sido capaz de captar la atención de las autoridades del Poder Ejecutivo, el Congreso y la ciudadanía en general, logrando sinergias entre sus distintas competencias (CLAD, 2012). El informe de García y García (2010) clasifica a Chile como el único país de América Latina y el Caribe que conjunta un alto grado de institucionalización del sistema con una importante utilización de sus resultados en la toma de decisiones (lo califican con 4.5 en una escala de 0 a 5, frente al 1.6 del promedio de América Latina y el Caribe).

Por otro lado, los esfuerzos también se han dado en materia de uso y difusión de las evaluaciones con la creación de incentivos individuales y colectivos (monetarios y de reconocimiento) que muestran la voluntad gubernamental de lograr que la evaluación sea vista como una herramienta para mejorar la gestión pública y no como una práctica obligada y punitiva.

También merece reconocimiento el impulso que se le ha dado a la capacitación y formación universitaria de especialistas en evaluación mediante la diversificación de la oferta académica y la creación de laboratorios de evaluación.

No obstante, existen tareas pendientes como la flexibilización de los procesos de asignación de recursos para modificar programas, el diseño de TR para realizar evaluaciones menos estandarizadas (predomina el enfoque de Marco Lógico), el énfasis en la medición de la satisfacción de los clientes o beneficiarios, y la realización de una evaluación de los resultados alcanzados por el propio SECG, entre otros.

México

México se incorpora de manera sistemática a los procesos de evaluación casi al finalizar el siglo XX. El número de sus trabajos crece aceleradamente y los tipos de evaluación se diversifican al reestructurarse el sistema con la creación a nivel federal del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) en 2006, al que en 2008 se agregaría otro órgano mucho más pequeño y de carácter local que aportó un enfoque alternativo a la evaluación: el Consejo de Evaluación del Desarrollo Social del Distrito Federal (Evalúa DF).

En esta parte del trabajo se desarrollan los temas presentados en los anteriores casos nacionales (su proceso de institucionalización y profesionalización) pero se agrega una revisión más detallada de algunos sectores de política de particular relevancia: desarrollo social, educación, salud, medio ambiente y desarrollo rural.

CAPÍTULO VIII
Institucionalización y profesionalización de la evaluación

Myriam Cardozo Brum

Nuevamente se inicia con una síntesis de lo sucedido en México en materia de antecedentes de la institucionalización de la evaluación durante el siglo pasado,¹ vinculados con la política de desarrollo social. Posteriormente se presentan las novedades recientes en ese sector, y se incluye una revisión exhaustiva de políticas aún no abordadas: educativa, ambiental y de salud.

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 global 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 de la NGP como consolidar una cultura administrativa que propiciara la eficiencia y la productividad en la gestión de las instituciones

¹ Para una revisión detallada del proceso, véase Cardozo (2012a).

gubernamentales, promover la evaluación de la calidad y la transparencia en la prestación de los servicios públicos y prevenir y combatir la corrupción.

El enfoque anterior 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, enfocándolo a la evaluación de resultados de la gestión pública pero éste sólo 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 las primeras y en 1997 se inició el proyecto de evaluación externa más ambicioso solicitado por el gobierno mexicano, el del Programa de Educación, Salud y Alimentación (Progesa), en el campo de la política social.

Institucionalización

Al iniciar el nuevo siglo el Congreso decidió comenzar a exigir la evaluación externa anual de los programas financiados con recursos fiscales, sujetos a Reglas de Operación (RO) y la Secretaría de Desarrollo Social (Sedesol) creó una Subsecretaría encargada de la evaluación de los programas que se encontraban bajo su responsabilidad. En 2002 se estableció el Sistema de Seguimiento de Metas Presidenciales (Simep), bajo la responsabilidad de la entonces Secretaría de la Contraloría y el Desarrollo Administrativo (Secodam) y para uso interno del Gobierno Federal y, al año siguiente, se aprobó la Ley General de Desarrollo Social (LGDS) que preveía la creación de un Consejo Nacional de Evaluación como organismo descentralizado de la Administración Pública Federal, que recién sería creado en 2006 como Coneval y se complementaría en 2007 con el Sistema de Evaluación del Desempeño (SED) de la SHCP, basado en la Ley de Presupuesto y Responsabilidad Hacendaria aprobada un año antes.

También en 2007 fueron publicados los nuevos “Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal” (en adelante Lineamientos) que sustituirían al Acuerdo que había estado vigente desde 2002 para normar esta actividad. Éstos serán analizados con detalle en la segunda parte del libro pero se anticipan los tipos de evaluaciones que definen: estratégicas y de programas federales. A su vez las segundas se componen de: *a*) evaluaciones de consistencia y resultados (ECR) referidas al diseño, desempeño global y resultados basados en el seguimiento de indicadores, *b*) evaluaciones de indicadores, que analizan su pertinencia y alcance para la medición de resultados, *c*) evaluaciones de procesos que revisan la eficacia y eficiencia de los procesos operativos y su contribución al mejoramiento de la gestión, *d*) evaluaciones de impacto, que miden cambios en los indicadores de resultados atribuibles a la gestión del programa, y *e*) evaluaciones específicas, que incluyen todas las no comprendidas en los tipos previos.

Ese mismo año se 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 (según lo establecido en la Ley Federal de Presupuesto y Responsabilidad Hacendaria de 2006), con lo cual su número aumentó en forma muy considerable (Cuadro 11).

La Ley de Fiscalización y Rendición de Cuentas de la Federación, aprobada en 2009, estableció la realización de auditorías al desempeño a cargo de la Auditoría Superior de la Federación (ASF). Se trata de una atribución no coordinada con el sistema de evaluaciones externas que se viene analizando.

En 2010, el Banco Interamericano de Desarrollo (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, frente a un promedio de 1.6 en la región de América Latina y el Caribe, con una diferencia aproximada de 1.3 respecto de Chile que ocupa el primer lugar), lo que indica que aún hay retos importantes que superar en materia de desempeño gubernamental.

CUADRO 11

México: evaluaciones programadas y realizadas por el Coneval 2007-2011 +

Tipo de evaluación	2007 P*	2007 R*	2008 P	2008 R	2009 P	2009 R	2010 P	2010 R	2011 P	2011 R	Total P	Total R	% R/P
Diseño	13	14	34	19	22	33	15	17	17	25	101	108	106.9
Consistencia y resultados	103	111	0	0	0	0	0	0	173	0***	276	111	40.2
Procesos	0	0	0	0	5	5	8	8	2	1	15	14	93.3
Específicas	6	6	18	18	27	22	3	1	0	0	54	47	87.0
Específicas de desempeño	0	0	133	131	131	132	126	133	0	0	390	396	101.5
Impacto	2	3	6	8	4	4	0	1	0	0	12	16	133.3
Estratégicas	2	2	1	1	4	1	3	1	2	0	12	5	41.7
Total	126	136	192	177	193	197	155	161	194	26	860	697	81.0
Complementarias**	-	0	-	5	-	5	-	4	-	4	-	18	-

+ Las cifras reportadas por las distintas fuentes presentan diferencias, se optó por presentar las de la SHCP por ser más completas y parecer más coherentes. Por ejemplo, Székely (2013) informa un total de 201 evaluaciones más para 2007-2011 (incluye 40 integrales que la SHCP no reporta), a las que suma 40 complementarias y 15 realizadas en el 2012 para obtener la suma total de 913.

Notas: * P es programado y R realizado.

** Las evaluaciones complementarias no se incluyen en el PAF.

*** Éstas aún no se habían terminado a la fecha en que se elaboró el cuadro (enero de 2012) pero Székely (2013) informa 140 realizadas.

Fuente: adaptado y corregido de SHCP (2012a).

El Coneval también ha sido objeto de evaluaciones y auditorías, como la del Banco Mundial (BM) en 2008 y la de la ASF de 2010. Más recientemente Székely (2013)² aplicó 24 entrevistas y encuestas a diversos actores relevantes en el sector de evaluación y medición de la pobreza en el mundo [académicos, funcionarios del BID, BM, Programa de las Naciones Unidas para el Desarrollo (PNUD), instituciones internacionales de evaluación de programas sociales y de los gobiernos de Perú y Sudáfrica] para conocer su percepción sobre Coneval. 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).

Székely también evaluó los resultados alcanzados por Coneval en comparación con su mandato original, en dos aspectos fundamentales (calidad de la información que genera y sustentabilidad), considerándolos de muy alto nivel (i.e. la estandarización de TR). En relación con el nivel de satisfacción respecto del Coneval, 48% de los funcionarios encuestados manifestó estar muy satisfecho y 39% algo satisfecho; mientras que en cuanto a su información y análisis, 56% y 31% los consideró respectivamente algo o muy requeridos para los elaboradores de políticas públicas y los tomadores de decisión. Un tercer aspecto (la utilización de la información generada) fue parcialmente cubierto por considerar que no se disponía de la información necesaria para realizar su valoración completa. El trabajo acaba con un ejercicio de análisis costo-beneficio del funcionamiento del Consejo Nacional de Evaluación de la Política de Desarrollo Social.

Otra evaluación (Rojas, 2013)³ se concentró en el desempeño organizacional del Coneval para pensar en su futuro desarrollo. 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, etcétera), 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

² Se trata de un documento muy completo de 218 páginas.

³ Consta de 80 páginas de extensión.

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.

Por otra parte, a partir de las evaluaciones del periodo 2011-2012, las dependencias y entidades de la administración pública 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, las que serán analizadas en la tercera parte de este libro. El propio Programa para un gobierno cercano y moderno 2013-2018 plantea que: “[...] en muchas ocasiones las evaluaciones y los ASM adoptados son vistos más como un trámite que deben atender las dependencias o entidades con un enfoque primordialmente procedimental, que de mejora al propio programa”. Por lo que no basta con tener evaluaciones sino que éstas deben ser de calidad y utilizadas para mejorar el diseño y el desempeño de los programas.

Además, el programa en cuestión 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. A continuación 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. De éstas se resaltan las más directamente vinculadas con los procesos de evaluación en estudio:

- Línea 1.1.6 “Incorporar la participación y contraloría ciudadana en el diseño, seguimiento y evaluación de las políticas públicas del Gobierno de la República”, a cargo de la Secretaría de Hacienda y Crédito Público (SHCP).
- Línea 1.5.6 “Fomentar el uso de los resultados del seguimiento y evaluación de los programas presupuestales como instrumento de contraloría ciudadana”, también bajo la responsabilidad de la Secretaría de Hacienda y Crédito Público.
- Líneas 2.1.2 “Incorporar los indicadores de los objetivos sectoriales al SED” por parte de la SHCP, y 2.1.5 “Consolidar el proceso de mejora de las MIR de programas presupuestales de forma que reflejen su contribución a las prioridades nacionales” bajo la responsabilidad de la SHCP, la Secretaría de

la Función Pública (SFP) y el Consejo Nacional de Evaluación de la Política de Desarrollo Social.

- Estrategias 2.2 “Vincular el SED con las asignaciones presupuestarias”, a cargo de la SHCP y la SFP; Estrategia 2.3 “Fortalecer el proceso de seguimiento y evaluación al desempeño de la APF”, responsabilidad de la SHCP, la SFP y el Coneval; y Estrategia 2.4 “Mejorar la calidad del gasto federalizado con base en los preceptos del SED”, nuevamente coordinadas por la SHCP y la Secretaría de la Función Pública.
- Línea 4.1.6 “Evaluar los resultados de los procesos a través de indicadores, para promover la mejora continua de su operación”, coordinada por la Secretaría de la Función Pública.

En algunas de las líneas y estrategias mencionadas, el 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. Esto se consumó con la reforma político-electoral, aprobada por el Congreso y publicada en el *Diario Oficial de la Federación* (DOF) el 10 de febrero de 2014, que le otorgó plena autonomía 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.

Al respecto no se conoce si los argumentos que habrían fundamentado el cambio coinciden con los manifestados por los funcionarios antes mencionados, pero su conveniencia ha sido discutida por el propio Consejo. Éste considera que, en los hechos, ha contado con autonomía desde su creación, a pesar de formar parte de la Sedesol, debido a múltiples causas como la capacidad de su personal, el compromiso y desempeño de sus consejeros, el carácter democrático de los gobiernos desde 2006, el respeto de los presidentes de la República y secretarios de desarrollo social durante este periodo, y muy especialmente, a la estructura interna que el Congreso le atribuyó. Se refiere fundamentalmente al proceso de elección por convocatoria pública de los integrantes de su Órgano de Gobierno, a los requisitos que éstos deben cumplir, a su extracción académica

(no funcionarios), que continúan trabajando en sus instituciones, que quedan al margen de la línea de mando de la Sedesol y conforman mayoría en el Coneval, así como a su desempeño *apartidista*.

El Coneval se planteaba que la reforma le daba más autonomía de derecho frente al Poder Ejecutivo, pero no fortalecería de igual forma su credibilidad, la que dependería de las leyes secundarias que se aprobaran. En consecuencia se ha planteado un conjunto de peligros: 1) puesto que los investigadores académicos serán elegidos en adelante por la Cámara de Diputados, su nombramiento podría *partidizarse* y el Coneval se volvería más dependiente del Poder Legislativo; 2) riesgo de que no se mantuvieran los requisitos académicos al aprobarse la reglamentación de la reforma; 3) que se nombraran nuevos consejeros antes que se dispusiera de las reglas mencionadas;⁴ 4) si los consejeros deberán trabajar de tiempo completo se aproximarían más a la Cámara para ser elegidos y a la realización de una carrera como funcionarios, perdiendo la riqueza y la independencia que les da la pertenencia a la academia, al tiempo que aumentarían sensiblemente los gastos de nómina del Consejo; 5) se teme también que reviva la búsqueda de errores y el énfasis punitivo de la evaluación, que podrían provocar el ocultamiento o “maquillaje” de la información; y 6) se sostenía la importancia de que los trabajadores del Coneval se mantuvieran dentro del Apartado A del artículo 123 constitucional (Coneval, 2014d).

Más allá de que no todos los especialistas en el tema están tan convencidos de la selección *apartidista* de los actuales consejeros o de que no convenga involucrar a funcionarios y políticos en su composición, ellos agregaban nuevas inquietudes: si las recomendaciones, que ahora se agregan expresamente a las atribuciones del Consejo, seguirán siendo de aplicación voluntaria o vinculante, si el énfasis de las evaluaciones seguirá concentrándose en los programas públicos o pasará a recaer en los lineamientos de políticas, si el Coneval también se hará cargo de las evaluaciones estatales y municipales o se limitará a expedir lineamientos para su desarrollo. También la necesidad de modificar leyes como la LGDS para armonizarla con el cambio y prever los mecanismos de coordinación entre el Consejo y otras dependencias gubernamentales, particularmente SFP

⁴ Los consejeros terminaban sus funciones el 31 de marzo de 2014 y la reforma establecía que los nuevos serían nombrados antes del 11 de abril, lo que no ha sucedido.

y SHCP. En general las dudas sobre las razones que llevaron a la Cámara a estas modificaciones y la composición y funcionamiento futuros del Coneval se comparten y provocan temor a un posible retroceso. También podría constituir la oportunidad de que se refuerce la utilización de las evaluaciones, uno de los puntos débiles más importantes de los trabajos realizados hasta ahora, y que se establezca una relación más adecuada con el Poder Legislativo y los gobiernos estatales y municipales. En particular, el foro convocado por el Centro de Investigación y Docencia Económicas (CIDE) en que se manifestaron estas inquietudes hubiera deseado conocer los puntos de vista del Congreso pero ningún diputado respondió a la invitación (CIDE, 2014).

En particular, Cejudo (2014) reconoce la reforma como una oportunidad para reforzar y mejorar lo que se venía haciendo mediante:

[...] una toma de decisiones más acertada en el diseño de los programas presupuestarios, que ayude a resolver los problemas de falta de coherencia entre políticas públicas, que vuelva realidad la promesa de que las asignaciones presupuestales toman en cuenta los hallazgos de las evaluaciones, que facilite la rendición de cuentas sobre la sustancia de la acción gubernamental, y que, en suma, lleve a mejores políticas, con mejores resultados para los ciudadanos.

Sezékly manifiesta su opinión:

[...] puede ser un elemento de fortaleza para el Coneval debido a que lo dessectoriza de la Sedesol, propiciando así un espectro más amplio de su accionar y una mayor independencia del Ejecutivo y del presupuesto gubernamental. A la luz de la historia de la institución desde sus orígenes, este cambio parecería como una evolución natural para consolidar su autoridad moral (2013:62).

En cambio considera riesgosa la nueva forma de nombrar a los consejeros y su duración, que podría pasar a ser dominada por criterios políticos en menoscabo de los principios básicos de fortaleza técnica y credenciales académicas que han prevalecido hasta ahora. También considera muy preocupante el otorgar poder de veto al presidente de la República, lo que pone en tela de juicio la independencia de los integrantes de la institución y aumenta el riesgo de subordinación de los Consejeros al gobierno en turno.

En medio del debate presentado, el 14 de octubre de 2014 la Comisión de Desarrollo Social aprobó por unanimidad el dictamen por el que se expide la Ley del Coneval y se reforman, derogan y adicionan diversas disposiciones de la LGDS para garantizar su plena autonomía. Los diputados de la instancia legislativa coincidieron en que es necesaria la separación del Coneval del sector coordinado por la Sedesol para que la valoración que realice de los programas, objetivos, metas y acciones de la política de desarrollo social sea más objetiva e imparcial, que las recomendaciones que emita tengan mayor fuerza para su cumplimiento y que el reto de los sistemas de evaluación sea lograr que los diagnósticos realizados por los evaluadores funcionen como una herramienta para identificar áreas de oportunidad y encauzarla en medidas de ajuste y corrección de las políticas, y no como una manera de reprobar a funcionarios públicos o programas de gobierno (Cámara de Diputados, 2014).

El pleno de la Cámara también lo aprobó casi por unanimidad⁵ el 28 de octubre de 2014. De esta forma el Coneval será también un órgano de consulta y asesoría para las dependencias y entidades de los tres niveles de gobierno de la Administración Pública, deberá emitir sus propias recomendaciones, y los resultados de sus evaluaciones se darán a conocer en el DOF, así como en su sitio de Internet y deberán ser entregados a las Comisiones de Desarrollo Social de las cámaras de Diputados y de Senadores, así como a la Sedesol. Contará con un comité técnico de carácter auxiliar consultivo y supeditado al Consejo General, el cual lo apoyará en la formulación de la metodología para la medición de la pobreza y la evaluación de la política de desarrollo social. Sus sesiones ordinarias serán de carácter público. El artículo séptimo transitorio establece que la integración del Consejo, la elección de los primeros consejeros y del presidente se realizará de manera simultánea por única ocasión, y dentro de los 60 días posteriores a la publicación del presente decreto (Cámara de Diputados, 29 de octubre de 2014). En adelante, los criterios para la elección de consejeros serán establecidos por las leyes secundarias que se aprueben.

La creación de un segundo Consejo levanta críticas y sospechas:

¿Qué nos quieren decir los legisladores con esta propuesta? ¿Que los integrantes del consejo principal serán más políticos que técnicos y por ello necesitan de

⁵ 414 votos a favor, tres abstenciones y uno en contra.

los segundos? ¿Son conscientes del impacto presupuestal de esta decisión? ¿Calibraron lo que esto implica para la gobernanza del nuevo Coneval? Además, en esta propuesta, la partidización de la selección sigue latente (Jaime, 2014).

Dejando a un lado esta nueva situación administrativa del Coneval es necesario mencionar que recientemente, han sido creados nuevos órganos federales encargados de la evaluación en temas educativos y ambientales que serán presentados en los apartados de políticas especializadas. 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.

Finalmente, además de las acciones mencionadas a nivel federal, en 2007 se decidió la creación de Evalúa DF, encargado de la evaluación de políticas y programas sociales a nivel de la capital (no incluye otros sectores), mismo que quedó integrado al siguiente año⁶ y que a la fecha ha realizado 28 trabajos de evaluación de políticas, programas y servicios públicos (Cuadro 12).

Además, en el marco del modelo de Gestión para Resultados, la Coordinación General de Modernización Administrativa del Gobierno del Distrito Federal (GDF) desarrolló el Programa de Monitoreo y Evaluación del Desempeño Gubernamental 2010-2012 (Promoeva) destinado a evaluar sus instituciones a partir de sistemas de indicadores, estadísticas y encuestas, y emitir recomendaciones basadas en información objetiva y confiable. Su información se concentró en el Sistema del mismo nombre (Sipromoeva).

Profesionalización de la evaluación

Asociaciones profesionales

México no ha contado con asociaciones profesionales independientes, salvo la Asociación Mexicana de Profesionales de la Evaluación (AMPE), que reúne principalmente a especialistas del desarrollo rural, por lo que no es representativa

⁶ Sus primeros trabajos se relatan y analizan en Cardozo (2012b).

CUADRO 12
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 la página web del Evalúa DF.

de todos los sectores ni tiene cobertura nacional y la incipiente Academia Nacional de Evaluación en México (Aceval) a la fecha ha organizado dos congresos, el tercero que tenía programado para el año pasado en Chihuahua debió cancelarlo. Por otra parte, muchos evaluadores participan en los congresos y publicaciones virtuales de la Redlacme, impulsada por el BM, el BID y Coneval, que cuenta con un capítulo mexicano (Red Mexicana de Monitoreo y Evaluación, Redmexme) y de la Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe (ReLac).

Publicaciones

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.⁷

⁷ Algunos de los libros publicados han sido: Mejía (2003), Magaña y Vargas (2004), Cardozo (2006), Vargas (2007), Castro (2007), Cortés *et al.* (2008), Cardozo (2009), Ramos *et al.* (2011), Cardozo (2012a y b), Cardozo y Mundo (2012), Arellano (2012), Ramos (2012), Ramos (2013) y Maldonado y Galíndez (2013).

Programas de formación

En materia de programas de formación, México no ha contado hasta ahora con programas académicos especializados en el tema, sólo algunos pocos cursos en posgrados más amplios.⁸ 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 existen programas especialmente enfocados al desarrollo de esta incipiente profesión⁹ y puede constatarse el reducido número de tesis vinculadas al tema, que en los últimos años se han venido incrementado. 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 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 (CEDRSSA) de la Cámara de Diputados y otro promovido por la Universidad Autónoma Metropolitana, Unidad 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 Distrito Federal. Por otra parte, Coneval junto con la Comisión Económica para América Latina (CEPAL) y, por otro lado, la SHCP, han estado capacitando intensamente en torno al diseño del marco lógico. La SHCP (2012) informa que de marzo 2009 a diciembre 2011 instrumentó un curso de educación a distancia en la Matriz de Indicadores para Resultados (MIR) que alcanzó a 2 459 servidores públicos.

⁸ Maestría en Políticas Públicas (UAM-Xochimilco), también existen Maestrías en Evaluación Educativa (UNAM) y en Evaluación y Administración de Proyectos (Universidad Panamericana).

⁹ El Instituto Tecnológico Autónomo de México (ITAM) es la única institución que ha contado con un centro especializado en materia de evaluación, pero se trata de un enfoque *ex-ante*, especialmente dirigido a proyecto de desarrollo regional.

Recientemente se ha formado también el Center for Learning on Evaluation and Result (CLEAR) con sede en el 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.¹⁰

Hace un año inició la impartición de su primer Diplomado de Políticas Públicas y Evaluación y actualmente está desarrollando el tercero, con un costo de 33 000.00 pesos mexicanos, bastante elevado para el contexto latinoamericano.

Con apoyo de CLEAR y la planta de investigadores del Centro de Investigación en Evaluación y Encuestas el INSP, también se aprobó en octubre de 2013 un programa virtual de especialidad de un año de duración en Evaluación Integral de Programas y Políticas de Desarrollo Social que debía iniciar en septiembre pasado. Pretende 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 decisiones, 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 que operen en los niveles local, estatal, nacional o internacional. 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 \$47 300.00.

Por último, el Coneval se muestra más preocupado por la capacitación continua de quienes están realizando las evaluaciones que por la formación en

¹⁰ [<http://www.clear-la.cide.edu/node/6>].

¹¹ Respuesta del Coneval al cuestionario diseñado para esta investigación.

general,¹¹ 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 pobreza 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 (Coneval, 2014a). A nivel municipal, las instituciones convocantes de la Conferencia Anual del Municipio (CAM) organizan para 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 SED, el que deberá estar listo a fines de 2015.

La exposición realizada muestra a México como un país que inició un poco tarde sus evaluaciones pero ha crecido aceleradamente y atraviesa un momento de incertidumbre relacionada con las posibles consecuencias de los cambios legales que estarán afectando el funcionamiento del Coneval. Queda por analizar la calidad y utilización de sus evaluaciones, lo que será realizado en el tercera parte de este libro. También requiere incrementar el número de sus programas de formación y especialmente ofrecerlos gratuitamente o a precios más accesibles, así como conformar asociaciones profesionales independientes y crear al menos una revista especializada en el tema. En los siguientes capítulos se profundizará en el estudio de algunos sectores de políticas y programas que presentan un nivel de avance importante en materia de evaluación.

CAPÍTULO IX

Desarrollo social

Myriam Cardozo Brum

Durante el siglo XX México desarrolló programas sociales nacionales a cargo de diversas secretarías de Estado (Salud, Educación Pública, Desarrollo Social, etcétera), organismos descentralizados y desconcentrados (Instituto Mexicano del Seguro Social, Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado, Instituto del Fondo Nacional de la Vivienda para los Trabajadores, Fondo de la Vivienda del 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, Leche Industrializada Conasupo, Distribuidora Conasupo, etcétera), entre otros. Muchos de estos programas se inspiraron en las propuestas del Estado de bienestar europeo.

En décadas recientes se pueden citar como los más relevantes: 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), el Progreso, emprendido por Ernesto Zedillo, y el Programa de Desarrollo Humano Oportunidades (PDHO), iniciado con Vicente Fox (2000-2006) y recientemente sustituido por Prospera (2014). A partir de la década de 1980, los programas que habían pretendido avanzar hacia

la universalización de sus beneficios fueron sustituidos por otros focalizados hacia la atención de la pobreza, especialmente la de carácter extremo.¹

Estos cambios en la política social federal fueron provocados por un recrudescimiento de la pobreza agudizado por la crisis de 1994-1995 y un giro de la política mexicana hacia el enfoque neoliberal. Entre 1996 y 2006 el Comité de Medición de la Pobreza informó de su disminución; sin embargo esta tendencia ha vuelto a revertirse y constituye la causa actualmente esgrimida por el Ejecutivo para introducir Prospera, agregando al desarrollo de capacidades un componente de generación de oportunidades de trabajo, especialmente mediante el desarrollo de proyectos productivos.

A nivel del Gobierno del Distrito Federal (GDF), éste ha defendido una política social diferente desde 1997, con la llegada del Partido de la Revolución Democrática (PRD). Dicha postura ha quedado plasmada en la Ley de Desarrollo Social (LDSDF) aprobada en el año 2000, mediante la definición de un conjunto de principios rectores que incluyen entre otros: la universalidad combinada con criterios de focalización territorial (no familiar o individual), la integralidad, el avance en el cumplimiento de los derechos económicos, sociales, culturales y ambientales, y la apertura de oportunidades de participación social (Cardozo, 2011).

Puesto 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 (Cardozo, 2012a), su especificidad debe ser tomada en cuenta para su evaluación.

La efectividad de un programa social depende del grado en que éste recupera y responde a las necesidades sociales, lo que implica disponer de un diagnóstico que priorice los problemas a atender (Rossi y Freeman, 1993), aplicar mecanismos de control y evaluación desde el inicio, y difundir públicamente la información generada para aumentar la transparencia y garantizar la participación social (Franco, 2003). Fortalecer el monitoreo y la evaluación de impacto de los programas sociales es también un reto mexicano identificado por el Banco Mundial (BM) (2004).

¹ Para profundizar en esta evolución, véase Cardozo (2012a).

Nuestro sistema de evaluación se ha concentrado desde su inicio en los programas sociales por lo que los antecedentes generales ya revisados y los específicos en este campo se encuentran imbricados. En 1990-1994 encontramos intentos por evaluar el Pronasol; sin embargo, el sistema inició sistemáticamente en 1998 con la evaluación de Progresá, y durante el 2001-2005 continuó especialmente con los programas a cargo de la Sedesol y la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Sagarpa). Como ya se mencionó, a partir de 2007 ha incorporado a los de otras secretarías. En 2012, de las 278 acciones y programas que componían el inventario del Coneval, 24 correspondían a Sedesol (8.6%) (Coneval, 2012a).

Entre las revisiones realizadas a programas sociales disponemos de las evaluaciones de Progresá 1998-2000, las auditorías de desempeño realizadas por la ASF en 1999-2001, y de una evaluación realizada en el marco de la campaña “El Banco Mundial en la mira de las mujeres”. Su resultados no resultan consistentes: mientras las primeras destacan una focalización más clara que en otros programas de América Latina, una importante sinergia al considerar simultáneamente los tres sectores clave (salud, educación y alimentación), fuerte impacto en el ingreso real de las familias, efectividad del condicionamiento de las transferencias, etcétera, la segunda insiste, entre otros, en la persistencia de brechas importantes en el desempeño de beneficiarios y el promedio de los estudiantes o la falta de un sistema adecuado de indicadores, y la tercera es mucho más negativa al valorar la exclusión de zonas indígenas lejanas, la falta de participación, el aumento del trabajo femenino no remunerado, etcétera. Al mismo tiempo que el BM refería evidencias de éxito y recomendaba la reproducción del programa en otros países latinoamericanos.

Las evaluaciones externas generalizadas a programas federales estuvieron coordinadas por la Subsecretaría de Prospectiva, Planeación y Evaluación de la Sedesol en 2001-2006. Fueron muy desiguales en calidad y costos (Cardozo, 2006). En 2002 se evaluaron 31 programas sociales, que en 2003 se redujeron a 16, de los cuales 11 formaban parte del grupo inicial.² En éstos se intentaba

² Desarrollo Humano Oportunidades, Abasto Social de Leche, Tortilla, Abasto Rural, Coinversión Social, Atención a Jornaleros Agrícolas, Empleo Temporal, Iniciativa Ciudadana 3X1, entre los principales programas de 2001 que continuaron vigentes en 2002.

cubrir todo el proceso seguido por los programas salvo el diseño (recursos, operación, cobertura y focalización, resultados e impactos) y se aplicaba el trabajo de campo como complemento del de gabinete. Las expectativas eran muy altas pero también la falta de experiencia y las dificultades del marco normativo y la misma organización de los procesos.

El periodo iniciado por el Coneval en 2007 ya fue mencionado en el capítulo anterior. Las evaluaciones dejaron de buscar la integralidad y se especializaron en alguno de los tipos mencionados por los Lineamientos de 2007.

CUADRO 13
Evaluaciones de programas de desarrollo social 2007-2014

Tipo de evaluación	2007	2008	2009	2010	2011	2012	2013	2014	Total
Consistencia y resultados	16				19				35
Diseño	5	3	1		1		1		11
Proceso		1	1			1	2		5
Impacto			1						1
Específica			1	1			12		14
Específica desempeño		20	19	19		26			84
Complementaria		1		3					4
Diagnósticos								7	7
Metaevaluaciones						20			20
Total	21	25	23	23	20	47	15	7	181

Fuente: elaboración propia con datos extraídos de las evaluaciones publicadas en el sitio web de la Sedesol y el Coneval, consultadas el 20 de octubre de 2014.

El Cuadro 13 muestra que casi la mitad de los trabajos realizados son Evaluaciones Específicas de Desempeño (EED), las más simples de las mencionadas, seguidas de ECR aplicadas sólo en dos años. También hay un buen número de lo que la Sedesol denominó metaevaluaciones a las que se hará referencia en la segunda parte del libro; su realización en 2012 casi duplicó el total de ese año. Resalta la casi ausencia de evaluaciones de impacto, la reciente aparición de las de diagnóstico y la falta de evaluaciones de indicadores previstas por los Lineamientos.

Las sucesivas metaevaluaciones realizadas de manera independiente a las evaluaciones federales (Cardozo, 2006, 2009, 2012a) han mostrado tanto avances como retos pendientes. Entre los primeros, se resalta un importante mejoramiento del marco normativo, una lenta 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 retos persisten algunos identificados desde el inicio de los procesos: contar con un fondo de recursos para la contratación de evaluadores, independiente 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, en la práctica federal se sigue cuestionando su conveniencia, mientras en otros países (países escandinavos, Suiza, Francia, Canadá, etcétera) constituye una estrategia ampliamente aceptada. Por ello se insiste en la necesidad de poner a prueba su utilidad en nuestro contexto mediante una estrategia incremental. Su ausencia podría interpretarse como indicadora de que el gobierno no está interesado en impulsar el aprendizaje democrático directo, contando con ciudadanos informados, críticos y propositivos, sino sólo por utilizar las evaluaciones para racionalizar el gasto social y ganar legitimidad (Cardozo, 2012a).

Por su parte, las evaluaciones del GDF se han venido realizando conforme a una práctica novedosa y emergente frente al modelo tradicional federal, basada en propuestas y procesos de planeación que implican distintos valores y criterios de organización y decisión. Resaltan como diferencias en la organización de los procesos: la preparación inicial del proceso mediante la impartición de cursos y diplomados en evaluación y la elaboración de una guía 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 transparentado en la convocatoria; las recomendaciones efectuadas por el Comité con carácter vinculatorio; y la propuesta de amplia difusión de resultados en periódicos y revistas, además de internet.

Asimismo se identifican diferencias en los métodos de evaluación empleados: evaluaciones construidas en función de las características específicas del programa (con TR *ad hoc*); reconocimiento de la existencia de diferentes puntos de vista; contribución tanto de los métodos cuantitativos como de los cualitativos; consideración de los estudios experimentales como 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; 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 (Cardozo, 2012b).

Cada Consejo, dentro de su enfoque, necesitaba mejorar la calidad desigual de sus trabajos de evaluación del desarrollo social y, a pesar de sus diferencias normativo-organizacionales, los dos enfrentaban el problema de la poca utilización de sus recomendaciones. La evolución reciente de estos dos aspectos será analizada en la segunda y tercera parte del libro respectivamente.

CAPÍTULO X

Educación

Talina Olvera Mejía

En México, la educación ha sido un tema de la agenda de los gobiernos en turno. Su cobertura fue en un inicio el principal problema, después el discurso giró hacia aspectos más cualitativos. En los últimos años de la década de 1970 la baja calidad se comenzó a identificar como un problema serio; actualmente existe un eje de las políticas educativas para enfrentarlo, junto con el de equidad. En este contexto, a partir de la década de 1990 se impulsan medidas para evaluar la calidad del sistema educativo, enfocadas principalmente en los resultados del aprendizaje, no en las políticas o programas. Sin embargo, los ámbitos de la evaluación educativa son más variados y pueden abarcar además de los logros de los alumnos (aprendizaje), la formación y el desempeño docente, los procesos educativos, los insumos, los actores, los programas de estudio, el funcionamiento del sistema, y las políticas y los programas educativos, entre otros.

En este capítulo se hace una revisión de lo realizado en materia de evaluación, focalizando el estudio hacia las políticas y los programas de educación básica y media superior.¹ Para ello, primero se realiza una breve descripción del proceso seguido por la evaluación educativa en México, para dar cuenta de la(s) dimensión(es) en que se ha enfocado esta actividad. Después se analizan

¹ Sólo se hace referencia a la educación básica y media superior dado que el Instituto Nacional para la Evaluación de la Educación en México se centra en estos niveles y sería imposible desarrollar también todo lo sucedido en materia de educación superior.

las atribuciones que en esta materia tienen dos organismos con competencia en evaluación: el Instituto Nacional para la Evaluación de la Educación (INEE), en tanto máxima autoridad en materia de evaluación educativa, y el Coneval, como instancia evaluadora de la política y programas sociales federales.

Calidad y evaluación educativa

La evaluación educativa ha estado enfocada en el aprendizaje, a partir de las pruebas estandarizadas como medio para medir sus resultados. Sin embargo, también es indispensable identificar y explicar si las políticas y los programas aplicados han permitido o no lograr los objetivos planteados, y es a partir de la evaluación que se puede valorar su impacto en la práctica educativa y contribuir a su mejora.

La experiencia evaluativa se remonta a la década de 1970 con la aplicación de las primeras pruebas a gran escala para el ingreso a la educación secundaria (1972) y con la creación (1974) de la Subdirección de Evaluación y Acreditación² adscrita a la Dirección General de Planeación Educativa de la Secretaría de Educación Pública (SEP), lo que favoreció el papel de la evaluación al menos a nivel estructural. En 1984 se estableció la Dirección General de Evaluación (DGE). Pero, los esfuerzos encaminados hacia la evaluación se caracterizaron por la descoordinación y la falta de seguimiento de las acciones, y por el poco o nulo aprovechamiento de los hallazgos.

En la segunda mitad de la década de 1990, como respuesta a esta desarticulación, se conformó un Sistema Nacional de Evaluación Educativa (SNEE), en torno de la DGE, para evaluar la calidad de los servicios educativos, principalmente el aprovechamiento de los alumnos y los factores que lo explican. Se evaluaron los programas compensatorios financiados por organismos internacionales pero sólo en su dimensión de aprendizaje, como el Programa para Abatir el Rezago Educativo (PARE),³ el Programa para Abatir el Rezago en

² Que después se convertiría en la Dirección General de Evaluación de Políticas (DGEP).

³ Primer programa compensatorio financiado con recursos del Banco Mundial, puesto en marcha en 1992. Su propósito era mejorar la calidad de la educación primaria en aquellas regiones indígenas, rurales y semiurbanas con altos índices de marginación social y económica.

Educación Básica (PAREB) y el Progreso. También la participación de México en pruebas internacionales y la reforma de la Ley General de Educación (LGE) de 1993 relativa a la evaluación del Sistema Educativo Nacional (SEN), entre otras.

Por su parte, la evaluación externa relativa a los programas educativos federales inició en 2001 en cumplimiento de la Ley de Presupuesto de Egresos de la Federación y en 2007 se sujetó a los Lineamientos emitidos por Coneval-SHCP-SFP para el conjunto de la Administración Pública Federal y al Sistema de Evaluación del Desempeño hacia un Presupuesto basado en Resultados. A nivel local, Evalúa DF ha realizado a la fecha tres evaluaciones externas en el campo educativo: de diseño y operación del Programa de Estímulos para el Bachillerato Universal en 2008, del diseño de los Servicios Educativos que proporciona el Instituto de Educación Media Superior del Distrito Federal en 2012, y de los resultados de la política de Inclusión en la Educación Media Superior del Gobierno del Distrito Federal, en 2013.

El INEE y su ámbito de acción en la evaluación educativa

En el ámbito educativo existía una necesidad de encontrar contrapesos adecuados a las tendencias centralistas de la SEP y a las corporativas del Sindicato Nacional de Trabajadores de la Educación (SNTE) mediante la creación de un órgano de evaluación independiente del gobierno. El equipo de transición educativa de Vicente Fox propuso la creación de un organismo especializado, técnico, con un grado considerable de autonomía que se encargara de realizar evaluaciones en gran escala del sistema educativo, actividad que, como se ha visto, realizaba exclusivamente la Dirección General de Evaluación.

El Programa Nacional de Educación 2001-2006 señaló a la consolidación del sistema de evaluación como una meta prioritaria de la política educativa. Pero debido a cuestiones técnicas y políticas el nacimiento del Instituto debió esperar más de un año mientras se discutía la mejor alternativa legal para crearlo: la posición de la Comisión de Educación del Senado en el sentido de hacerlo mediante la aprobación de una ley que le otorgara plena autonomía a pesar de que requería efectuar negociaciones para alcanzar la mayoría necesaria (ninguno de los grupos parlamentarios la detentaba) o la expedición de un

decreto presidencial que no le podría atribuir plena autonomía pero tampoco implicaba hacer negociaciones (Martínez y Blanco, 2010).

La segunda opción fue la que prevaleció y el presidente de México expidió el decreto de creación del INEE el 8 de agosto de 2002, con lo que éste sólo adquirió el carácter de organismo descentralizado de la SEP, que mantuvo su atribución exclusiva para evaluar al SEN y fijar los lineamientos generales. El INEE funcionó como órgano de consulta y apoyo técnico de la SEP. Su objetivo fue ofrecer a las autoridades educativas las herramientas adecuadas para realizar la evaluación de los diferentes elementos del sistema educativo como medio para mejorar la calidad educativa, poniendo énfasis en la difusión oportuna y transparente de los resultados.⁴ Respecto al aprendizaje, realizó evaluaciones integrales para conocer el rendimiento de los estudiantes, los factores de contexto que lo explican y la calidad del sistema educativo, para apoyar la toma de decisiones y la rendición de cuentas a la sociedad.

El INEE tenía incidencia en cuatro ámbitos vinculados con la calidad de: las estadísticas y los indicadores preparados por la SEP, las pruebas nacionales⁵ la coordinación de las evaluaciones internacionales⁶ y las evaluaciones de las escuelas (recursos y procesos). Como se puede ver, ninguna de estas atribuciones se vinculaba con la evaluación de programas educativos federales.

El 15 de mayo de 2012, el presidente de la República firmó la reforma al decreto de creación del INEE que entró en vigor el 15 de agosto y supuso un cambio de estatus jurídico, un ajuste organizacional y el ejercicio de nuevas funciones, operando como un organismo descentralizado, no sectorizado con personalidad jurídica y patrimonio propio. Esta condición suprimió la supervisión que le realizaba la SEP, pero mantuvo su papel de apoyo y órgano consultor en la evaluación del SEN. Sus cuatro funciones principales fueron la generación y desarrollo de información, estadísticas e indicadores educativos; el diseño y desarrollo de instrumentos de medición y evaluación; el desarrollo y

⁴ El INEE tuvo un papel relevante en la difusión de evaluaciones a inicios de la década pasada, ya sea en sitios web o en publicaciones.

⁵ Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) y Exámenes para la Calidad y el Logro Educativos (Excale). La primera aplicación de Excale se realizó en junio de 2005 y de ENLACE en junio de 2006.

⁶ Programa Internacional para la Evaluación de Estudiantes (PISA por sus siglas en inglés).

promoción de investigación en evaluación educativa; y la difusión y vinculación para el uso de la información y el conocimiento.

En febrero de 2013 se procedió a la reforma constitucional mediante la cual se modificó el artículo 3 para incluir la fracción IX que dice: “El Instituto Nacional para la Evaluación de la Educación será un organismo público autónomo, con personalidad jurídica y patrimonio propio”. De esta forma, el INEE se convirtió en un órgano constitucionalmente autónomo⁷ y en la máxima autoridad en materia de evaluación educativa a nivel nacional. Otro paso importante para concretar su autonomía fue la publicación en septiembre de 2013 de la Ley Reglamentaria de la fracción IX del artículo 3 constitucional (Ley del INEE). Este marco jurídico se completó con la publicación de su Estatuto Orgánico en octubre de 2013, que definió el nuevo diseño institucional, estableciendo de manera precisa las facultades y funciones de sus órganos de gobierno, órganos colegiados y unidades administrativas.

Como consecuencia de las anteriores modificaciones, el objeto del INEE quedó establecido como coordinar el SNEE⁸ y evaluar la calidad, el desempeño y los resultados del SEN⁹ en educación básica y media superior, pública y privada,

⁷ En México existe una escala de seis niveles de autonomía; en ésta la “autonomía constitucional” es un atributo deseable debido a que permite, en teoría, disponer de libertad en la toma de decisiones; atender con mayor objetividad e imparcialidad los asuntos de su competencia; y emitir recomendaciones ajenas a intereses particulares, sindicales o políticos.

⁸ El SNEE lo constituye el propio INEE y las autoridades educativas estatal y federal. Para su funcionamiento cuenta con la Conferencia, cuyo propósito es intercambiar información y experiencias relativas a la evaluación educativa. El SNEE también está constituido por los componentes, procesos y resultados de la evaluación; por los parámetros e indicadores educativos; por los lineamientos y las directrices de la evaluación; por los procedimientos de difusión de los resultados de las evaluaciones; y por los mecanismos, procedimientos e instrumentos de coordinación destinados al funcionamiento del SNEE. Su objetivo es contribuir a garantizar que los servicios educativos sean de calidad.

⁹ El SEN está formado por los educandos, educadores y los padres de familia; las autoridades educativas; el Servicio Profesional Docente; los planes, programas, métodos y material educativos; las instituciones educativas del Estado y sus organismos descentralizados; las instituciones de los particulares con autorización o con reconocimiento de validez oficial de estudios; las instituciones de educación superior a las que la ley otorga autonomía; la evaluación educativa; el Sistema de Información y Gestión Educativa; y la infraestructura educativa (artículo 10 LGE).

en todas sus modalidades¹⁰ (artículo 25 de la Ley del INEE y artículo 1 del Estatuto del INEE). De manera más precisa se le atribuye el diseño y realización de mediciones y evaluaciones a las instituciones, políticas y programas educativos. También pasó a ser el encargado de expedir los lineamientos a los que se sujetarán las autoridades educativas federal y locales para llevar a cabo las evaluaciones que les correspondan, mismas que serán obligatorias; así como también de diseñar, poner en marcha y mantener actualizado el sistema de indicadores educativos y de resultados de las evaluaciones. En este conjunto de atribuciones aparece finalmente incluida la evaluación de políticas y programas educativos.

La Ley del INEE refiere en su artículo 6 que la evaluación “consiste en la acción de emitir juicios de valor que resultan de comparar los resultados de una medición u observación de componentes, procesos o resultados del Sistema Educativo Nacional con un referente previamente establecido”. Esta definición, que se limita a la evaluación del SEN, deja fuera a las políticas y programa educativos federales. Complementariamente, el artículo 7 establece que la evaluación del SEN tiene como fin contribuir a mejorar la calidad de la educación y a la formulación de políticas educativas, además del diseño e implementación de los planes y programas que de éstas se deriven, con lo que excluye su evaluación.

El artículo 8 establece que la evaluación será obligatoria y periódica, aunque no especifica sus tiempos. Tampoco define el alcance de las evaluaciones, es decir, los aspectos mínimos que deberán incluir o los criterios de valoración que tomarán en cuenta. Agrega que la evaluación será sistemática e integral, lo que indica que no será exclusivamente de resultados. Pero el foco de la evaluación del INEE se ha centrado en los maestros,¹¹ descuidando la evaluación de otros

¹⁰ Si bien las instituciones de educación superior a las que la LGE les otorga autonomía constituyen también parte del sistema educativo nacional, como ya se mencionó, el INEE sólo evalúa la educación básica y media superior. Sin embargo, estas instituciones podrán suscribir convenios con el Instituto en los términos señalados en la Ley del Instituto Nacional para la Evaluación de la Educación.

¹¹ Al maestro se le considera muchas veces como el único factor explicativo de lo que ocurre en la educación. Las atribuciones del INEE tienen que ver con la expedición de los lineamientos a que se sujetarán las autoridades educativas para llevar a cabo las funciones de evaluación que les correspondan para el ingreso, promoción, reconocimiento y permanencia en el servicio profesional docente en la educación obligatoria (artículo 28, Ley del INEE).

factores importantes en los resultados educativos. Si bien la normatividad induce a pensar en la facultad del INEE para evaluar políticas y programas educativos al plantear en el artículo 25 de su Ley que “el Instituto diseñará y realizará mediciones y evaluaciones que correspondan a componentes, procesos o resultados del Sistema Educativo Nacional respecto a los atributos de educandos, docentes y autoridades escolares, así como, de las características de instituciones, políticas y programas educativos”, es la misma normatividad la que provoca ambigüedad al respecto. Los artículos 6 y 7 parecieran contradecir lo anterior, cuando limitan la evaluación al SEN, mismo que entre los elementos que lo constituyen no comprende a las políticas y programas federales.

Puesto que una evaluación no cumple su función si sus resultados y recomendaciones no se ponen en práctica, es importante saber si las expedidas por el INEE serán sólo observaciones de nivel técnico, no vinculatorias, o si las partes interesadas se comprometerán a cumplirlas y, en caso de no hacerlo recibirían alguna sanción. Pero la Ley no establece su obligatoriedad, sólo señala que las autoridades e instituciones educativas deberán hacer pública su respuesta en relación a ellas en un plazo no mayor a 60 días naturales (artículo 51 de la Ley del INEE) y que no hacerlo es causa de responsabilidad administrativa (artículo 65 de la Ley del INEE). Así que tanto la LGE como la normatividad del INEE no establecen el carácter vinculatorio de las recomendaciones, y al referirse a éstas lo hacen de manera ambigua, con términos como “atendiendo a las directrices”, “promover”, “directrices que sean relevantes para contribuir”. Incluso los fines del INEE se redactan en términos de acciones como “contribuir” y “fomentar”. Por esto, el uso de las evaluaciones y la vinculación de las recomendaciones sigue siendo una tarea pendiente para el Instituto. Lo importante es que la SEP tome decisiones a partir de los resultados que se deriven de las evaluaciones, y que el INEE no se limite a ser sólo un órgano asesor, sin capacidad de influir realmente en la política educativa del país.

Después de décadas de ocultamiento de la información, hoy se establece que se deben dar a conocer los resultados sobre el desarrollo y avance de la educación (artículo 31 de la LGE), además del informe anual que el INEE debe presentar al Congreso de la Unión sobre el estado que guarda el SEN derivado de sus evaluaciones, así como las directrices que éste emita y los acuerdos de su Junta, para que la ciudadanía pueda acceder a la información relacionada con el Sistema Nacional de Evaluación Educativa.

La Junta de Gobierno es el órgano superior de dirección del INEE; en su composición se ha decidido emular la experiencia del Coneval: que sus integrantes tengan un perfil académico (no político) para evitar que criterios políticos (de partidos o gobiernos) obstaculicen las valoraciones educativas, repercutiendo en la calidad de la educación. Pero también hay que tener cuidado en no caer en la formación de un grupo de expertos, alejado del contacto e intereses de los actores educativos, para lo cual el INEE cuenta con dos órganos: el Consejo Social Consultivo de Evaluación de la Educación, que tiene la función de conocer, opinar, y dar seguimiento a los resultados de las evaluaciones y directrices que realice el Instituto, por lo que sus acuerdos no son vinculantes, y los Consejos Técnicos Especializados para coadyuvar en el cumplimiento de las atribuciones del INEE a partir del asesoramiento técnico y metodológico en materia de evaluación educativa (artículo 47 del Estatuto del INEE).

Un asunto interesante que no se aborda en el marco jurídico, pero que sería importante aclarar es si el INEE podría realizar consultas locales en foros abiertos de discusión o reuniones con los actores educativos para tratar temas relevantes en cuanto a la evaluación y calidad educativa de cada región del país, de manera de involucrar a un número mayor de actores interesados. Habría que poner atención también en la influencia efectiva que dichos resultados tendrían en las decisiones y acciones del INEE, para evitar que el Consejo se limitara a comunicar sus propuestas y escuchar opiniones sin ninguna consecuencia.

En conclusión, las evaluaciones del INEE, cuyos resultados están publicados en su sitio web, han estado enfocadas en las pruebas estandarizadas y el aprendizaje de español, matemáticas, lectura y escritura; en las pruebas internacionales, en las prácticas pedagógicas y docentes, en los aprendizajes en el aula, en los factores asociados con el aprendizaje y en la infraestructura escolar. No obstante que las evaluaciones de programas están previstas en su normatividad (artículo 25 de la Ley del INEE) y de haber publicado un folleto en el que sus autores sostienen la necesidad de efectuarlas (Anderson *et al.*, 2006), el Instituto no ha realizado evaluaciones de políticas o programas, ni se vislumbra que a corto plazo incursione en ese ámbito.

El Coneval y la evaluación de programas federales¹²

Antes de 2006, algunos programas educativos innovadores a los que se les destinaba una importante cantidad de fondos públicos¹³ fueron sujetos a evaluaciones dentro del marco de transparencia y rendición de cuentas sobre el uso de los recursos públicos establecido por la Cámara de Diputados para ser considerado en la elaboración del Presupuesto de Egresos de la Federación (PEF). En estos casos, la SEP encargaba su realización a organismos externos, especialmente instituciones educativas de prestigio y sus informes eran publicados en los respectivos sitios web de los programas.

Si bien las evaluaciones a políticas y programas educativos no han constituido un ámbito de acción del INEE, a partir de 2007 las referidas especialmente a los programas federales han estado a cargo del Coneval y sujetas a su normatividad. Una justificación de esta atribución para evaluar los programas educativos es que éstos y las acciones federales de desarrollo social buscan contribuir a la preservación de los derechos sociales y al bienestar económico de los ciudadanos. Tan sólo en el Inventario de Programas y Acciones Federales de Desarrollo Social del Coneval de 2012, que contó con 278 programas y acciones, el mayor número estuvo dirigido al derecho a la educación: 103 intervenciones (37%) y 92 de ellas fueron operadas por la Secretaría de Educación Pública.

A partir del año mencionado, en el portal electrónico de la SEP se publican las evaluaciones realizadas a sus programas. Éstas abarcan todos los niveles educativos (básico, medio superior y superior), así como los programas de sus organismos descentralizados (Consejo Nacional de Fomento Educativo, Conafe; Comisión Nacional de Cultura Física y Deporte, Conade; Instituto Nacional para la Educación de los Adultos, INEA; Comisión Nacional de Libros de Texto Gratuitos, Conaliteg) y desconcentrados (Consejo Nacional para la Cultura y las Artes, Conaculta). En la mayoría de los casos la Dirección General de Evaluación de Políticas, y en el resto, la Unidad de Planeación y Evaluación de Políticas Públicas de la SEP dan seguimiento a las evaluaciones.

¹² Debido a que la naturaleza y atribuciones de este organismo ya fueron abordadas, nos enfocamos sólo a la injerencia que tiene en evaluación de programas educativos.

¹³ Un ejemplo es el Programa Escuelas de Calidad, que desde su inicio en el ciclo escolar 2001-2002 contempló este tipo de evaluaciones por parte de la Secretaría de Educación Pública.

También en el caso de educación, la normatividad aplicada por el Coneval a todos los programas federales define los tipos de evaluación que se pueden realizar y las características de los equipos de evaluadores. Aunque en 2007 se registraron sólo ECR y de diseño, en años posteriores se han llevado a cabo todos los tipos de evaluaciones externas previstos en los Lineamientos, salvo el de indicadores. En el Cuadro 14 se presentan las evaluaciones realizadas de 2007 al 2014 en el sector, que han sido publicadas en el sitio en internet de la Secretaría de Educación Pública.

CUADRO 14
Evaluaciones de programas de educación 2007-2014

Tipo de evaluación	2007	2008	2009	2010	2011	2012	2013	2014	Total
Consistencia y resultados	24	1	3		35				63
Diseño	4	3	8	1	11	3*	1		31
Proceso					1	1			2
Impacto						1			1
Específica						43*	58*		101
Específica desempeño		35	37	36		41			149
Complementaria		8							8
Diagnósticos								14	14
Total	28	47	48	37	47	89	59	14	369

* Estas evaluaciones sólo aparecen en el catálogo del Coneval.

Fuente: elaboración propia con datos extraídos de las evaluaciones publicadas en el sitio web de la SEP y el Coneval, consultadas el 3 de octubre de 2014.

El cuadro muestra que el mayor número de evaluaciones se aplicó en 2012. También una mayor concentración en evaluaciones específicas y EED. Estas últimas responden a los objetivos de gestión por resultados al valorar el desempeño de los programas durante el ejercicio fiscal mediante un simple trabajo de gabinete. Las evaluaciones de diseño se muestran con continuidad porque son obligatorias para los nuevos programas de cada año. A pesar de que las ECR ocupan el tercer lugar por el número de evaluaciones realizadas no han tenido una continuidad en su realización a lo largo de los años. Llama la atención las pocas evaluaciones de proceso e impacto realizadas en estos siete

años, ambas son esenciales para tomar decisiones tanto en la mejora de los programas como del problema que se pretende abatir. Los datos muestran que hace falta mayor trabajo de campo y de análisis en las evaluaciones más que reportes que reproducen información año con año y no arrojan datos relevantes para cumplir con las funciones básicas de toda evaluación.

En resumen, podemos ubicar dos vertientes de la evaluación en educación: la primera como un procedimiento para evaluar el aprendizaje necesario en los estudiantes, la segunda como instrumento para valorar los componentes del sistema desde la perspectiva de una política pública o programa. En México, la evaluación en el ámbito educativo ha estado más relacionada con la primera vertiente en que las pruebas estandarizadas se han convertido en el referente para formular juicios sobre la calidad educativa. Sin embargo, la evaluación debe ir más allá y conocer los procesos que han generado esos resultados y los factores que los han condicionado, pues la evaluación educativa vinculada solamente a aprendizajes no es útil si no se complementa con los demás elementos del sistema. Y aunque tradicionalmente la evaluación de políticas y programas educativos ha sido menos visible y menos debatida que la de aprendizajes, la primera genera conocimiento sobre la problemática educativa evaluada y pone a prueba las hipótesis del programa. En la práctica, el INEE tiene la responsabilidad de evaluar la primera vertiente, mientras que el Coneval se encarga de la segunda.

CAPÍTULO XI

Medio ambiente

*Esthela Sotelo Núñez**
*Angélica Rosas Huerta***
*Laura Rodríguez Cardozo****

Desde hace varios años el gobierno mexicano se ha dado a la tarea de impulsar un conjunto de acciones orientadas a mejorar la calidad ambiental, proteger su capital natural y cooperar internacionalmente para hacer frente a los retos climáticos compartidos con el resto del planeta. Durante el siglo pasado priorizó la creación de Áreas Naturales Protegidas (ANP) y en las últimas décadas ha fortalecido su política ambiental por medio de instituciones renovadas, impulso de acciones e instrumentos para la conservación de la biodiversidad y los recursos naturales, así como en la gestión de los recursos hídricos, la promoción del desarrollo sustentable y la ampliación de la agenda ambiental a temas como el cambio climático. Se cuenta con evaluaciones que valoran los logros de los programas en casos como la gestión de calidad del aire y de residuos sólidos, la conservación y el uso sustentable de los bosques, la biodiversidad y la gestión del cambio climático.

En este capítulo se presenta un panorama general de las evaluaciones en el sector ambiental en México y, en especial, de las realizadas a las ANP y a los

* Doctora en Estudios Urbanos y Ambientales. Profesora-investigadora del Departamento de Política y Cultura, UAM-Xochimilco, México.

** Doctora en Ciencias Sociales. Profesora-investigadora del Departamento de Política y Cultura, UAM-Xochimilco, México.

*** Doctora en Medio Ambiente y Desarrollo, Ensenada, Baja California, México.

programas de cambio climático. El objetivo es ofrecer al lector una mirada rápida de las características inherentes al sector ambiental, que imponen retos particulares en términos evaluativos.

La evaluación de políticas y programas ambientales

La evaluación de políticas y programas ambientales supone un enorme reto para los gobiernos que deciden institucionalizar su práctica. Esto es así debido a que, aunado a las dificultades inherentes a la práctica evaluativa en general, muchas de las áreas o temas que conforman la agenda ambiental precisan de indicadores de resultados que sólo son observables en horizontes temporales de muy largo plazo.¹ A diario, tomadores de decisión e investigadores de todo el mundo enfrentan el problema de la falta de datos disponibles, la no comparabilidad entre éstos y, en algunos casos, la complejidad de interpretación de los mismos.

Los problemas de falta de datos para verificar resultados, y la no-comparabilidad de los mismos, han sido reconocidos como el principal reto a enfrentar dentro del campo de la evaluación ambiental en diversos países. Al respecto, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y el Programa de las Naciones Unidas para el Desarrollo (PNUD) han emprendido importantes esfuerzos para que instituciones tales como la Agencia Europea del Medio Ambiente (AEMA) destinen gran cantidad de recursos a la construcción de un sistema de recolección y reporte de datos científicos más sistemático, coherente e internacionalmente comparable, como insumo para evaluar el desempeño y los resultados de sus programas ambientales. En diferentes países, instituciones internacionales como el Panel Intergubernamental de Cambio Climático (PICC) colaboran activamente con agencias nacionales en el diseño de formatos universales para reportar los avances de los programas ambientales. Incluso en algunos países de la Unión Europea, los datos construidos con estos criterios estandarizados se usan como insumos de las evaluaciones a programas

¹ Por ejemplo, el suelo es un recurso no renovable a escala humana, por lo que aquellos programas presupuestarios que tengan como objetivo rehabilitar y regenerar dicho recurso van a enfrentar dificultades para monitorear el logro de sus objetivos.

ambientales (Crabbé y Leroy, 2008). A pesar de estos esfuerzos, la AEMA ha afirmado que en muchas ocasiones es imposible llevar a cabo evaluaciones exhaustivas de los programas ambientales debido, entre otras razones, a que la Unión Europea y sus Estados miembro ponen muy poca atención al tema (AEMA, 2001a).

En México la práctica evaluativa en el sector ambiental se ha vuelto un ejercicio regular desde el año 2000. A pesar de ello, la información disponible sobre las evaluaciones es aún limitada y dispersa, lo que dificulta dimensionar el nivel de cumplimiento de los objetivos sectoriales en este tema. De acuerdo con un informe sobre el desempeño ambiental en México, presentado por la OCDE en 2013, las evaluaciones que se han realizado a los programas de este sector “[...] aún no tienen un enfoque lo suficientemente centrado en resultados ambientales”; los indicadores disponibles permiten sólo el seguimiento de avances en la implementación de actividades (OCDE, 2013:56).

Si bien se han realizado sistemáticamente evaluaciones externas a los programas del sector ambiental, únicamente están disponibles para consulta electrónica las realizadas a partir de 2007² debido a que su publicación se volvió obligatoria a partir de la reglamentación de la Ley General del Desarrollo Social en 2006. Ese mismo año se creó el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) y las evaluaciones externas de los programas federales ambientales comenzaron a normarse por los Lineamientos, ya referidos en capítulos anteriores³ y, en consecuencia, a clasificarse en evaluaciones estratégicas y evaluaciones de programas. Estas últimas pueden ser de consistencia y resultados, de indicadores, de procesos, de impacto y específicas. Además, ante la ausencia de evaluaciones de diseño en el periodo 2000-2006, Coneval introdujo como un elemento básico de los procesos de evaluación la elaboración de la MIR de cada programa, en el marco del SED y del Programa de Mejora de la Gestión de la Secretaría de Medio Ambiente y Recursos Naturales (Semarnat, 2012a).

² [<http://www.semarnat.gob.mx/programas/evaluaciones>], fecha de consulta: 9 de octubre de 2014.

³ Aunque en este sector también existen evaluaciones internas y mixtas (o parcialmente contratadas), tales como el PNANP, y algunas evaluaciones realizadas por la Conagua en colaboración con diferentes centros de investigación.

En este contexto, en 2010, se llevaron a cabo 97 evaluaciones de 37 programas presupuestarios de la Semarnat. En 2011 se realizó un ejercicio de monitoreo a partir de 51 MIR; en conjuntos éstas contienen 693 indicadores que se clasifican en estratégicos y de gestión. De acuerdo con el citado informe de la OCDE, en los últimos años han empezado a surgir evaluaciones cuyos resultados aportan información valiosa, que redundan en beneficios ambientales. Tal es el caso de la evaluación de los programas de gestión para mejorar la calidad del aire (ProAire) del Valle de México (OCDE, 2013:55).

Del conjunto de acciones y programas que componían el inventario del Coneval en 2012, a Semarnat correspondían 30 de un total de 278 (10.8%). En el periodo 2007-2012 las evaluaciones realizadas en este ramo se concentraron en programas a cargo de los organismos sectorizados; es decir, la Comisión Nacional del Agua (Conagua), la Comisión Nacional Forestal (Conafor), la Comisión Nacional de Áreas Naturales Protegidas (Conanp) y el Instituto Nacional de Ecología y Cambio Climático (INECC). También hubo algunas evaluaciones al interior de la cabeza de sector, en aquellas áreas encargadas de la integración sectorial y regional, así como de la gestión de actividades riesgosas en la Dirección General de Política Ambiental e Integración Regional y Sectorial (DGPAIRS), la Subsecretaría de Planeación y Política Ambiental (SPPA) y, en menor medida, la Dirección General de Gestión Integral de Materiales y Actividades Riesgosas (DGGIMAR).

El tipo de evaluación más atendido para los programas ambientales en el periodo 2007-2012 fue la EED (40.15%), seguido de la ECR (23.36%). El tipo de evaluación menos considerado fue la de procesos y la evaluación específica con 0.73%, realizadas sólo en 2009 y 2012, respectivamente (Cuadro 15).

El programa más evaluado en el periodo 2007-2012 fue ProÁrbol (40 evaluaciones), con sus distintas vertientes, perteneciente a la Conafor. Le sigue el Programa de Conservación para el Desarrollo Sostenible (Procodes, antes Proders), implementado por la Conanp (6 evaluaciones). En los dos primeros años del sexenio 2013-2018 se han realizado 36 evaluaciones externas en el sector, siendo las más comunes las evaluaciones específicas de desempeño realizadas en 2013 (Cuadro 16).

CUADRO 15
Evaluaciones de programas de medio ambiente por tipo 2007-2012

Tipo de evaluación	2007	2008	2009	2010	2011	2012	Total
Consistencia y resultados	14					18	32
Específica						8	8
Diseño		1	7	6	13	3	30
Complementaria		9		1	4		14
Específica de desempeño		22	15	19			56
Procesos			1				1
Impacto	1						1
Total	15	32	23	26	17	29	142

Fuente: elaboración propia con base en los datos presentados por la cabeza de sector [<http://www.semarnat.gob.mx/programas/evaluaciones/anteriores>], fecha de consulta: 13 de enero de 2015.

Hasta octubre de 2014, los diversos programas ambientales de todo el sector cuentan con 163 evaluaciones externas, siendo 2008 el año con mayor número de éstas (32). No se han realizado aún evaluaciones de impacto ni de indicadores como tal, pero se ha procedido al diseño y rediseño de la Matriz de Indicadores para Resultados.

CUADRO 16
Evaluaciones de programas de medio ambiente 2013-2014

Tipo de evaluación	2013	2014	Total
Diseño	11		11
Complementarias	4	3	7
Específicas de desempeño	17		17
Consistencia y resultados		1	1
Total	32	4	36

Fuente: elaboración propia con base en los datos presentados por la cabeza de sector [<http://www.semarnat.gob.mx/programas/evaluaciones>], fecha de consulta: 13 de enero de 2015.

Hasta 2014, los diversos programas ambientales de todo el sector cuentan con 163 evaluaciones externas, siendo 2008 el año con mayor número de éstas (32). No se han realizado aún evaluaciones de impacto ni de indicadores como tal, pero se ha procedido al diseño y rediseño de las MIR. A continuación destacamos dos casos relevantes en la materia: evaluación en ANP y en programas de cambio climático.

Áreas naturales protegidas

Las ANP constituyen “porciones terrestres o acuáticas del territorio nacional representativas de los diversos ecosistemas, en donde el ambiente original no ha sido esencialmente alterado y que producen beneficios ecológicos cada vez más reconocidos y valorados” (Conanp, s/f:1). Han sido reconocidas internacionalmente como la principal herramienta para la conservación de especies y de ecosistemas, ya que proveen bienes y servicios esenciales para el uso sustentable de los recursos naturales (Programa de Naciones Unidas para el Medio Ambiente, PNUMA, s/f). Su institucionalización inicia en Europa en el siglo XIX y la mayoría de los países, incluido México, ha ido desarrollando sistemas extensivos de áreas protegidas. En América Latina, los esfuerzos para la administración y el manejo de áreas protegidas fueron muy escasos y desiguales hasta 1992 en que se realizó la Conferencia de las Naciones Unidas sobre el Medio Ambiente y el Desarrollo en Río de Janeiro, conocida como Cumbre de la Tierra.

Las ANP mexicanas (como instrumentos de política forestal e hídrica), creadas por decreto presidencial, se establecieron formalmente desde 1876 en que fue protegido el bosque Desierto de los Leones, en los alrededores de la Ciudad de México, pero es hasta la publicación de la Constitución Política de 1917 que se incluyeron regulaciones y limitaciones para el aprovechamiento de los recursos naturales (Conanp, s/f:1). A partir de la década de 1970 la conservación dejó de ser exclusiva del agua y los bosques, y se enfocó de manera más consistente, por un lado, a la biodiversidad en general y, por otro, a la incorporación expresa de las comunidades humanas a través de las Reservas de la Biósfera. En 1988, la publicación de la Ley General de Equilibrio Ecológico y de Protección

al Ambiente (LGEEPA) incorporó el desarrollo sustentable a las prioridades existentes en las áreas naturales protegidas.

A principios de la década de 1990 se produjo la institucionalización de la gestión ambiental en lo general y la de conservación ecológica en lo particular, con la creación de importantes organizaciones: en 1992 la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (Conabio), en 1994 la Secretaría de Medio Ambiente, Recursos Naturales y Pesca (Semarnap), en 1997 el Fondo para Áreas Naturales Protegidas (FANP),⁴ en 1998 el Sistema Nacional de Áreas Protegidas conocido como SINAP I, en el año 2000 la Conanp, y en 2002 el SINAP II.

Como consecuencia del proceso anterior, la creación y ampliación de ANP adquirió un impulso sin precedentes. En el periodo 1991-2011, el área bajo protección se incrementó del 5.4 al 12.2% de la superficie terrestre (UICN y UNEP-WCMC, 2012) para el 2011 había 174 ANP decretadas. Hoy en día, México cuenta con 176 ANP que cubren casi 13% (25 394 779 hectáreas) del territorio nacional y está comprometido con el Plan Estratégico para la Diversidad Biológica 2011-2020 y las metas de Aichi, Japón, en las que se estableció que, al menos el 17% de las zonas terrestres y de aguas continentales, y el 10% de las zonas marinas y costeras, se conservarán por medio de sistemas de áreas naturales protegidas.

La evolución brevemente referida muestra que las ANP han sido consideradas como un valioso instrumento de la política ambiental para la conservación y el desarrollo sustentable, por lo que es de suma importancia evaluar su buen funcionamiento desde el punto de vista ambiental y también en materia de desarrollo económico y social.

A nivel internacional, los primeros trabajos tendientes a la evaluación de ANP corresponden al inicio del presente siglo y en su mayoría se trata de propuestas metodológicas más que de ejercicios empíricos efectivamente aplicados; es el caso de Cifuentes *et al.* (2000), Hockings *et al.* (2000), y Kleiman *et al.* (2000), que resaltan la necesidad de evaluar todo el proceso y su efectividad, tanto en lo biológico como lo social. Posteriormente, De la Maza *et al.* (2003) analizaron

⁴ Esquema público-privado entre el Fondo Mexicano para la Conservación de la Naturaleza (FMCN), y la Comisión Nacional de Áreas Naturales Protegidas.

la evolución de las ANP de América Latina y el Caribe en 1992-2000. Otros trabajos relevantes fueron los de Steam *et al.* (2005), Leverington *et al.* (2008), y diversos artículos publicados en revistas como *New Directions for Evaluation* o *Journal of Evaluation and Program Planning*. Se destaca también la publicación por la Unión Internacional para la Conservación de la Naturaleza (UICN) de un manual para la evaluación de la efectividad, con base en indicadores biofísicos, socioeconómicos y de gobernabilidad (Pomeroy *et al.*, 2004), los que posteriormente fueron retomados en otros trabajos como en el caso del complejo Malindi-Watamu en Kenia (Muthiga, 2009).⁵

Si bien las ANP mexicanas iniciaron desde 1876, sus evaluaciones son escasas. En el 2001, la Conanp creó el Sistema de Información, Monitoreo y Evaluación para la Conservación (SIMEC), al que se integraron datos geográficos, indicadores y estadísticas bajo la metodología de Presión-Estado-Respuesta, con el propósito de evaluar los cambios y la condición ambiental de las ANP, a fin de contar con diagnósticos de las causas y efectos potenciales de los problemas presentes, así como predecir los futuros impactos de las actividades humanas en las mismas. Su objetivo ha sido “establecer un sistema que incorpore indicadores biológicos, geográficos, sociales y económicos que permitan analizar la efectividad e impacto en la aplicación de políticas públicas en las Regiones Prioritarias para la Conservación” (Conanp, s/f:1).

El SIMEC ha sido organizado para operar en tres ejes: información, monitoreo y evaluación. Cada uno de éstos puede visualizarse como un subsistema que implica actividades diferentes, pero que al final intenta la integración de sus componentes en un sistema único. Los tres ejes fueron establecidos para determinar la efectividad en la instrumentación de los programas institucionales y su impacto en la conservación de las áreas naturales protegidas.

En 2004 se planeó una evaluación de mediano término al Programa de Trabajo 2001-2006 (PT), encargado de coordinar las acciones en las ANP; esta evaluación no está disponible actualmente,⁶ pero con base en la información ofrecida por el Programa Nacional de Áreas Naturales Protegidas (PNANP)

⁵ Para una revisión más amplia a nivel internacional, véase Rodríguez y Espejel (en prensa).

⁶ Información confirmada en la entrevista a los funcionarios de la Dirección de Evaluación y Seguimiento.

que le dio continuidad en 2007-2012, se puede determinar que no se trató de una verdadera evaluación que analizara información sobre el estado de las ANP, menos aún en relación con el impacto logrado por el PT, sino de un ejercicio sobre la pertinencia y valoración de los 53 indicadores presentados por dicho programa (evaluación de indicadores).

Por otra parte, Figueroa y Sánchez-Cordero (2008) hicieron una evaluación de la efectividad de algunas ANP para prevenir el cambio de uso de suelo, considerándolo como la mayor causa de otros procesos de degradación, y encontraron que aún hay una porción significativa de ANP que son poco efectivas o no efectivas. Sin embargo, como ellos mismos lo mencionan, hicieron sólo una evaluación parcial o un diagnóstico preliminar.

En 2009, la Conanp volvió a planear una evaluación de mediano término; esta vez del PNANP. Pero nuevamente se trató de un ejercicio de valoración y rediseño de indicadores y de replaneación de metas para 2010-2012, sin mostrar ninguna información de las ANP, en un documento de extensión muy reducida.

En 2011 la Conanp elaboró 42 estudios sobre la valoración de la pérdida de ecosistemas naturales, que incluyeron 29 ANP decretadas entre 1937 y 2010, encontrando que de las 16 ANP que tuvieron únicamente un estudio, en 12 (75%) la velocidad de transformación resultó negativa y sólo en 4 (25%) fue positiva; mientras sólo 23.1% de las que tuvieron dos estudios pasaron de transformación negativa a positiva (ASF, 2011).

En el marco del sistema de evaluaciones externas del Coneval, las relativas a ANP han iniciado en 2011 con una evaluación de diseño del Programa de Vigilancia Comunitaria en Áreas Naturales Protegidas y Zonas de Influencia (Provicom); en 2012 al programa anterior se le aplicó un EED y se agregaron dos evaluaciones de diseño y dos específicas a los programas de Manejo de Áreas Naturales Protegidas (Promanp) y de Monitoreo Biológico en Áreas Naturales Protegidas (Promovi); en 2013 se aplicaron evaluaciones específicas a los tres programas mencionados; y en 2014 aún no se publica ninguna. Sorprende la ausencia del programa rector en el tema hasta 2012 (Pronanp).

Adicionalmente al trabajo realizado por la Administración Pública Federal que se relató, la Auditoría Superior de la Federación, órgano técnico de la Cámara de Diputados, realiza auditorías de desempeño de los programas públicos. En particular, en 2011 analizó la política ambiental en diversos rubros, incluida la biodiversidad y más concretamente las ANP. En esta última revisó la evolución

del territorio protegido comparándola con las metas internacionales asumidas en el Plan Estratégico para la Diversidad Biológica 2011-2020 de Aichi, Japón, y los estudios realizados por la Conanp que se acaban de referir.

La mayoría de los esfuerzos realizados miden principalmente el estatus de las poblaciones de fauna y flora y la calidad o salud de los ecosistemas, pero la revisión realizada no evidencia que los procesos económicos y sociales sean tomados en cuenta. No se identifican avances a nivel estatal.

Cambio climático

El estudio y comprensión del cambio climático como un problema global de origen humano, caracterizado por el aumento de la temperatura media global del aire, la variación de los regímenes hídricos y el aumento de eventos climáticos extremos, se ha desarrollado desde hace varias décadas. Sin embargo, fue terminando la década de 1980 cuando a escala internacional se institucionalizó no sólo como objeto de estudio, sino como ámbito de intervención pública.

Para contribuir a su atención se han desarrollado dos estrategias generales básicas: la adaptación y la mitigación. La adaptación orientada en articular mecanismos de respuesta ante los cambios, así como acciones dirigidas a reducir la vulnerabilidad ante los efectos adversos del cambio climático. En cambio, la mitigación busca reducir las emisiones de gases de efecto invernadero a partir de acciones que permitan modificar procedimientos, usar otras tecnologías, limpias y eficientes, así como prácticas ecológicas, sea para compensar o para evitar más emisiones de carbono.

En la década de 1990 y a principios del año 2000 se puso énfasis a la mitigación como respuesta al cambio climático (Parry *et al.*, 1998), centrada en el cumplimiento de los compromisos adquiridos internacionalmente. En la última década se ha incrementado la atención en la estrategia de mitigación (Pielke *et al.*, 2007). Actualmente, varios gobiernos nacionales como estatales han promovido un conjunto de acciones enmarcadas en ambas estrategias.

En el caso de México, la gestión climática se ha extendido en la última década hacia los tres niveles de gobierno; gradualmente los gobiernos federal, estatal y municipal han reconocido su responsabilidad en la materia y han diseñado y adoptado acciones específicas de mitigación y de adaptación al cambio climático.

La Estrategia Nacional de Cambio Climático (ENCC), dada a conocer en 2007, representa el primer instrumento de planeación que rige y orienta la política de cambio climático. En 2009 se publicó el Programa Especial de Cambio Climático 2009-2012 (PECC, 2009). Actualmente el gobierno mexicano cuenta con la segunda ENCC, dada a conocer en 2013, así como con el segundo PECC, que abarca el periodo de 2014-2018.

A nivel federal, el principal instrumento es la Ley General de Cambio Climático aprobada en 2012 y reformada en 2014 (LGCC, 2014) que establece, entre otros aspectos, que la labor de las autoridades gubernamentales no sólo se debe enfocar a formular, conducir y publicar la Estrategia Nacional y el Programa de Cambio Climático, sino también a llevar a cabo su instrumentación, seguimiento y evaluación (artículo 7). Con respecto a la evaluación de la política nacional de cambio climático, ésta podrá realizarse por la Coordinación de Evaluación⁷ o, mediante uno o varios organismos independientes. Esta evaluación debe ser periódica y sistemática para proponer, en su caso, su modificación, adición, o reorientación total o parcialmente (artículos 25 y 98).

A nivel estatal el problema del cambio climático también se ha atendido, actualmente existen diez entidades federativas que han puesto en práctica un Programa Estatal de Acción ante el Cambio Climático, mientras que otras dieciséis entidades lo están diseñando (INECC, 2014). En estos esfuerzos destaca la Ciudad de México debido a que fue el primer gobierno local mexicano que reconoció al cambio climático como un asunto de interés público y lo incorporó en su agenda; también ha sido el primero en nuestro país en contar con un programa específico para este problema.

Desde inicios de la década pasada el gobierno de la Ciudad ha desarrollado y fortalecido su marco legal, programático e institucional que le ha permitido llevar a cabo acciones para enfrentar el cambio climático, incluso antes que en el gobierno federal. En 2004 dio a conocer su Estrategia Local de Acción Climática y hace unos meses su Estrategia de Acción Climática para el periodo

⁷ Cabe destacar que la Coordinación de Evaluación se integrará por el titular del INECC y seis consejeros sociales, representantes de la comunidad científica, académica, técnica e industrial, con amplia experiencia en materia de medio ambiente, particularmente en temas relacionados con el cambio climático (artículo 23).

2014-2020. También ejecutó el Programa de Acción Climática 2008 a 2012 y ahora ha iniciado el segundo, de 2014 a 2020. En el 2011 se publicó la Ley de Mitigación y Adaptación al Cambio Climático y Desarrollo Sustentable para el Distrito Federal, en la cual también se considera a la evaluación como práctica constante dentro de la gestión del cambio climático.

En cuanto a la evaluación de los programas de cambio climático, a nivel federal existen estudios realizados por las propias dependencias gubernamentales que informan sobre los avances de las acciones presentadas en el Programa Especial de Cambio Climático 2009-2012 (Semarnat, 2012b). También se han realizado evaluaciones por parte de organismos internacionales en torno al desempeño ambiental; es el caso de la evaluación realizada por la OCDE en 2013.⁸

Además se cuenta con una evaluación del Programa Especial de Cambio Climático 2009-2012,⁹ que busca valorar el nivel de confianza de los avances reportados en mitigación, verificar la trazabilidad de la información¹⁰ en las metas de mitigación y adaptación y evaluar su impacto en las políticas públicas de mitigación a nivel federal. El alcance de la evaluación incluyó también la formulación de recomendaciones específicas para disminuir los rangos de incertidumbre y aumentar la trazabilidad en la implementación y seguimiento del Programa (IMCO, 2013).

⁸ Para una revisión detallada de los objetivos y resultados de la evaluación véase OCDE-Centro Mario Molina (2013).

⁹ Fue financiada por GIZ GmbH, organismo de cooperación al desarrollo del gobierno alemán, mediante fondos de la Iniciativa Internacional del Clima (ICI, por sus siglas en inglés) y realizada por DetNorske Veritas y el IMCO. DetNorske Veritas es una empresa global de manejo de riesgo, salvaguarda de la seguridad de la vida relacionada a temas de propiedad y medio ambiente. La empresa es autónoma e independiente y su visión es tener impacto global para asegurar la seguridad ambiental de sus clientes y de la sociedad en su conjunto. El IMCO es un centro de investigación independiente, sin fines de lucro y apolítico. El principal objetivo de la institución es mejorar la competitividad de México a partir de propuestas de política pública y análisis de cuantitativo robusto que permita la toma de decisiones.

¹⁰ La trazabilidad de la información se refiere a la existencia de documentos con los recursos, datos y las metodologías empleadas para estimar los avances y logros de cada una de las medidas de mitigación, adaptación, comunicación y educación contenidas en el programa (Centro Mario Molina, 2012:68).

En 2013 México firmó también un acuerdo con Francia para evaluar las políticas sobre el cambio climático. Este acuerdo incluyó el cofinanciamiento de 250 mil euros para el proyecto de evaluación a desarrollarse en dos años, bajo el liderazgo del INECC (Embajada de Francia en México, 2014).

A nivel estatal, el GDF también ha realizado trabajos de seguimiento y evaluación de su Programa de Acción Climática (PACCM) 2008-2012. Hasta antes del 2011 no se contaba con evaluaciones que permitieran identificar y valorar sus efectos, sólo existían algunos estudios realizados por el propio gobierno de la Ciudad de México que informaban sobre los avances de sus acciones y proyectos.¹¹ Fue hasta finales de 2012 cuando el Centro Mario Molina (CMM), a solicitud de la Sedema, elaboró una evaluación del PACCM, cuyo objetivo fue realizar una revisión del desempeño de los proyectos y acciones que lo integran y generar recomendaciones que permitieran fortalecer la contribución de la Ciudad de México a la reducción de emisiones de gases de efecto invernadero (GEI), así como mejorar su resiliencia a los impactos del cambio climático (CMM, 2012).

En el marco de dicha evaluación, el CMM publicó la *Guía metodológica para evaluar los programas de acción climática de gobiernos locales*, en la cual se presenta un conjunto de procedimientos y criterios para caracterizar la integridad de la estructura, contenido, gestión y difusión de un programa en la materia, en particular desarrolla procedimientos lógicos y secuenciales para analizar la trazabilidad del diseño, reporte, monitoreo y verificación de las medidas de mitigación en un programa local de cambio climático. Sin embargo, no incorpora métodos que permitan evaluar la ejecución de un programa de acción climática y estimar su efectividad.

Por último, en el marco de las evaluaciones realizadas por el Coneval sólo se localizaron dos, lo que muestra la casi ausencia de evaluaciones federales vinculadas al tema. Se trata de la evaluación de diseño al Programa de Mitigación y Adaptación al Cambio Climático de 2011 y la de diagnóstico a la Investigación en Cambio Climático, Sustentabilidad Ambiental y Crecimiento Verde de 2014.

¹¹ Para un mayor conocimiento sobre los avances en el programa véase GDF (2011).

Conclusiones

En el caso de las evaluaciones en ANP, a la dificultad inherente de medir las condiciones de conservación o deterioro de los elementos del medio físico, se suma el reto de integrar a los procesos económicos y sociales como parte integral del desarrollo sustentable de estas áreas, que hasta ahora no han sido considerados en las evaluaciones realizadas a esta estrategia.

En cuanto a los programas de cambio climático, las evaluaciones realizadas tanto por organizaciones nacionales (CMM) como internacionales (OCDE y GIZ) contribuyen principalmente a aportar desarrollos metodológicos para evaluar el diseño de la política de cambio climático. Sin embargo cada evaluación tiene sus propias particularidades: mientras la OCDE y el GIZ han enfocado sus esfuerzos evaluativos principalmente en robustecer y homologar la información disponible, la aportación del CMM se orienta a los procesos metodológicos que garantizan una información completa y confiable para medir la trazabilidad de las acciones y estimar la reducción de las emisiones de GEI. En consecuencia, el objetivo de realizar evaluaciones de operación, resultados e impacto de los programas sigue siendo un reto no superado.

Sin embargo, en los tres últimos sexenios la institucionalización de la evaluación en el sector ambiental mexicano ha avanzado considerablemente. Los temas que concentran la mayor cantidad de evaluaciones constituyen la columna vertebral de la gestión ambiental en México: *a*) gestión y distribución de recursos (en este caso agua y recursos forestales); *b*) conservación de los ecosistemas terrestres y costeros, a través de la protección del territorio; *c*) adaptación y mitigación al cambio climático; *d*) integración regional a partir de instrumentos de política tales como el Ordenamiento Ecológico (atribución de DGPAIRS), y *e*) prevención/control de la contaminación.

Esta cobertura temática da cuenta de que la agenda evaluativa dentro del sector ambiental se concentra en las áreas de interés prioritarias, por lo que podríamos suponer que los esfuerzos evaluativos se canalizan de forma eficiente, al menos en cuanto a la decantación temática. Sin embargo, los resultados asociados a esta práctica aún son difíciles de precisar, debido entre otras cosas a la dispersión de los datos y la no comparabilidad de los mismos, factores que dificultan llevar a cabo evaluaciones cruciales como las de impacto.

En el marco del Coneval existe un número importante de evaluaciones ambientales, pero no de procesos ni de impactos cuyos aportes son insustituibles, y tampoco en materia de cambio climático y ANP. La nueva Coordinación de Evaluación del INECC, compuesta con base en la experiencia de organizaciones similares (Coneval, Evalúa DF e INEE), será la responsable de recuperarlas en el futuro.

CAPÍTULO XII

Salud

Myriam Cardozo Brum

La salud, al igual que la educación, ha sido siempre un tema obligado de las agendas gubernamentales, en parte debido a los problemas de desigualdad y pobreza que afectan las condiciones de salud de la población, que resultan agravadas por el también desigual acceso a servicios.

Las políticas y los programas mexicanos de salud son encabezados a nivel federal por la Secretaría de Salud (SS), que también se encarga de atender a la población abierta (sin seguridad social), conjuntamente con los Servicios Estatales de Salud (Sesa)¹ y el Programa Instituto Mexicano del Seguro Social (IMSS)-Prospera. Pero debido a los recursos de que disponen, existen dos instituciones importantes de seguridad social que tienen programas propios y prestan servicios a los trabajadores del sector formal de la economía: el IMSS y el Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado (ISSSTE). Otra institución relevante es el Sistema Nacional para el Desarrollo Integral de la Familia (DIF). El conjunto de instituciones públicas y privadas compone el llamado Sistema Nacional de Salud, establecido después de la reforma a la Ley General de Salud de 1984.

El sector ha recuperado siempre las sugerencias emitidas a nivel internacional para encauzar el rumbo hacia los problemas más relevantes a nivel mundial (Organización Mundial de la Salud, OMS) y regional (Organización Paname-

¹ A los que se descentralizaron parte de las funciones del sector en el periodo 1985-2000.

ricana de la Salud, OPS). Originalmente la mayor preocupación de estos organismos fue la extensión de la cobertura; para lograrla México puso en marcha el Programa de Ampliación de Cobertura en Salud (PAC) que entre 1996 y 2000, ofreció un paquete básico de 12 intervenciones a una población de 10 millones de mexicanos que carecía de acceso regular a esos servicios, lo que implicaba la eliminación o reducción de las barreras de acceso (geográficas, sociales, culturales y económicas) a los mismos. En paralelo al PAC, en la segunda mitad de los años noventa, el IMSS inició el programa de Seguro de Salud para la Familia que permitió a los integrantes del sector informal de la economía con capacidad de pago inscribirse en el mismo mediante una cuota de acceso que varía en función de su nivel de ingresos. Aunque la cobertura ha estado creciendo, mantiene niveles inaceptables: según datos del Coneval, en 2008 había 44.8 millones de mexicanos sin acceso a servicios de salud, los que en 2010 disminuyeron a 35.8 millones (22.2% menos).

Posteriormente a los programas de cobertura, la atención se centró en los programas de vacunación. En años recientes a estas preocupaciones se han sumado otras como los costos de los servicios, la distribución del financiamiento público-privado, la calidad e incluso la calidez del trato. En relación con la calidad, ésta ha sido siempre muy desigual (mucho mejor en la seguridad social que en la SS) y fue descuidada al término del siglo pasado, lo que llevó a que los entrevistados de la Encuesta Nacional de Satisfacción de Servicios de Salud, realizada en el año 2000 por la Fundación Mexicana para la Salud, afirmaran que la calidad respecto de 1994 había empeorado (Gómez *et al.*, 2004).

Los programas que han intentado responder a los problemas previos son muchos y se clasifican en función de criterios diversos: por padecimientos (hipertensión arterial, cáncer, diabetes mellitus, etcétera), por grupos de edad de la población (materno-infantil, infancia, adolescencia, envejecimiento), por grupos sociales (población indígena), por niveles de atención (primaria, secundaria y terciaria), por su objetivo (curativo, de prevención y de promoción), entre otros. Además, existen programas intersectoriales que presentan un componente de salud como el actual Prospera y sus antecesores Oportunidades y Progresa.

En el sexenio 2000-2006 es importante subrayar la creación del Sistema de Protección Social en torno del Programa del Seguro Popular, con el objetivo de que las familias sin seguridad social tuvieran acceso a las acciones de protección en salud. Durante el sexenio 2006-2012 destaca el inicio del programa

Caravanas de la Salud, dirigido a brindar servicios a la población que se encuentra alejada de centros de salud, y del Programa del Sistema Integral de Calidad en Salud (Sicalidad). En 2012, el sector salud incluía 41 acciones y programas del total federal de 278 (15%) que integraban el Inventario de Programas y Acciones Federales de Desarrollo Social. La SS era responsable directa de 20 (Coneval, 2012a).

El Programa Sectorial de Salud 2013-2018 cuenta con seis objetivos: 1) consolidar las acciones de protección, promoción de la salud y prevención de enfermedades; 2) asegurar el acceso efectivo a servicios de salud con calidad; 3) reducir los riesgos que afectan la salud de la población en cualquier actividad de su vida; 4) cerrar las brechas existentes en salud entre diferentes grupos sociales y regiones del país; 5) asegurar la generación y el uso efectivo de los recursos en salud; y 6) avanzar en la construcción del Sistema Nacional de Salud Universal.

La evaluación también ha modificado sus objetivos, a la par de las políticas y programas: en las décadas de 1950 a 1970 la evaluación de los programas y servicios se concentró fundamentalmente en la medición de la eficacia (logro de metas) y la eficiencia (uso de los recursos/producción de servicios). Unas décadas más tarde estas modalidades fueron sustituidas por la evaluación simultánea de la estructura, los procesos y los resultados de los programas. Al finalizar la década de 1990 la evaluación comenzó a dirigirse al desempeño global de los sistemas y es resultado de una combinación de la aplicación de las políticas, la marcha de los programas y los resultados de las acciones de salud (Garrido y López, 2011).

Previo a la creación del Coneval en 2006 no fue posible encontrar un registro sistemático de las evaluaciones realizadas a las políticas y los programas de salud, pero éstas existieron como lo ejemplifica la evaluación de impacto de la reforma mexicana de salud 2001-2006 (Gakidou *et al.*, 2005). También se realizaron auditorías de desempeño por parte de la ASF como las efectuadas a los componentes de salud y alimentación del Progresá (2000 y 2001 respectivamente).

Actualmente, la evaluación de los programas de salud es responsabilidad de la Dirección General de Evaluación del Desempeño (DGED) de la SS, que participa en la elaboración del PAE junto con el Coneval, la SHCP y la SFP, y aporta información, supervisa la calidad, el cumplimiento de la normatividad y el buen desarrollo de todas las etapas de la evaluación, incluyendo la definición

de los ASM y su seguimiento. A diferencia de los sectores de Educación y Medio Ambiente, Salud no cuenta con un órgano de evaluación especializado en el tema, a pesar de su importancia.

En el caso de los programas prioritarios y especiales como el Sistema de Protección Social en Salud, dada las amplias expectativas que genera para el cumplimiento de los objetivos sociales en materia de salud a nivel nacional, se busca obtener evidencias sobre su instrumentación, ejecución y resultados, por lo que su evaluación queda específicamente a cargo de la Dirección de Evaluación de Programas Prioritarios y Especiales de Salud (DEPPES) de la DGED, que también coordina sus evaluaciones externas complementarias.

En el Cuadro 17 se presenta el número de evaluaciones externas realizadas en el sector desde la creación del Coneval. Resaltan muchas evaluaciones complementarias (29) y pocas ECR (9), en comparación con la distribución del total realizado por Coneval. Se han concentrado en programas de creación relativamente reciente considerados prioritarios: (Caravanas de la Salud, Sicalidad, Seguro Popular) cuyas evaluaciones se repiten a lo largo del cuadro en la modalidad de diseño, EED, ECR, complementarias y específicas. Además, son las únicas disponibles en la página de la Secretaría de Salud.

CUADRO 17
Evaluaciones de programas de salud 2007-2014

Tipo de evaluación	2007	2008	2009	2010	2011	2012	2013	2014	Total
Consistencia y resultados	2				7				9
Diseño	2	1	3	3	1				10
Proceso						1			1
Impacto		1							1
Específica						20	28		48
Específica desempeño		16	28	26		9			79
Complementaria	9	4	3	3	3	5	2		29
Diagnósticos								4	4
Integral			1	2	1	1			5
Estratégica			1						1
Total	13	22	36	34	12	36	30	4	187

Fuente: elaboración propia con datos extraídos de las evaluaciones publicadas en los portales electrónicos de la Secretaría de Salud y el Coneval, consultadas el 3 de diciembre de 2014.

Otros programas evaluados anualmente desde 2007 son Comunidades Saludables, Protección y Desarrollo Integral de la Infancia y Atención a Personas con Discapacidad (DIF), cuyas fichas ofrecen información desde 2008, al igual que Prevención y Atención contra las Adicciones, Reducción de Enfermedades Prevenibles por Vacunación y Reducción de la Mortalidad Materna. Ante el crecimiento del problema de la obesidad, desde 2009 se evalúa también el Programa de Prevención contra la Obesidad. Los programas del IMSS comenzaron a evaluarse en 2009 y los del ISSSTE en 2012.

En el caso del GDF también existen algunos programas de salud relevantes para la capital que han sido evaluados por Evalúa DF, como el diseño y operación de Acceso Gratuito a los Servicios Médicos a las Personas Residentes en el Distrito Federal que carecen de Seguridad Social Laboral y el de Atención Integral al Cáncer de Mama, ambos en 2008, y la operación, resultados e impacto de los Servicios de Salud Materno-Infantil del GDF 2006-2010, realizada en 2011. Además de las evaluaciones mencionadas, el sector salud ha contado históricamente con un detallado seguimiento de metas, a partir de un sistema de indicadores, con el fin de ajustar su operación para mejorar su desempeño. Por ejemplo, a nivel federal, la DGED dio seguimiento a las metas vinculadas con las 10 estrategias y las 74 líneas de acción del Programa Nacional de Salud 2007-2012, y actualmente debe monitorear los 16 indicadores de resultados reconocidos como prioritarios para el seguimiento periódico de los objetivos y metas establecidas en el nuevo Programa Sectorial.

En conclusión, el sector salud tiene una larga trayectoria en la elaboración de estadísticas que dan seguimiento a un sistema de indicadores ampliamente aceptado y desde 2007 cuenta con un promedio de 26 evaluaciones externas anuales coordinadas con el Coneval en 2007-2013. Aunque esa cifra corresponde aproximadamente a la mitad de las realizadas en el sector educativo, es similar a la alcanzada en Desarrollo Social, y dada la importancia del sector, habría que valorar la conveniencia de crear su propio órgano autónomo de evaluación.

CAPÍTULO XIII

Desarrollo rural

*Osmar Medina Urzúa**

A mediados de la década de 1980 fue desmantelado el cuerpo de extensionistas¹ que trabajaban como técnicos de la entonces llamada Secretaría de Agricultura y Recursos Hidráulicos. En este periodo, con la emisión del Plan Nacional de Desarrollo, la Secretaría asumió la coordinación del Programa de Desarrollo Rural. El acontecimiento es relevante, ya que dio pie, en 1988, a la configuración de los Distritos de Desarrollo Rural (DDR) y los Centro de Apoyo al Desarrollo Rural (Cader) mediante los cuales se organizó la acción e implementación de la política agropecuaria, pesquera, alimentaria y de desarrollo rural.

Actualmente en el sector rural de México habitan 26 millones de personas (13.1 millones de mujeres y 12.9 millones de hombres), 61.6% de ellas en condición de pobreza y 21.5 en pobreza extrema (Censo de Población y Vivienda, 2010). Además, ocupa a 6.7 millones de personas, de los cuales 86% se dedica a la actividad agrícola produciendo en predios con una baja escala productiva (menores a cinco hectáreas), altamente dispersos y en su mayoría de subsistencia (Sagarpa, 2013).

Los datos anteriores muestran una conexión intrínseca del sector primario y el ámbito rural en México, ligados a lo largo del tiempo en la política de

* Maestro en gobierno y asuntos públicos y jefe del Departamento de Control y Seguimiento, Secretaría de Gobernación, México.

¹ Sistema de servicios profesionales privados pagados con recursos públicos (Amsda, 2011).

desarrollo. Sin embargo, hasta décadas recientes no se tenía explicitado lo que debería entenderse por desarrollo rural en los documentos guía de la política sectorial nacional.

Fue en 1973, con la implementación del Programa de Inversiones para el Desarrollo Económico Rural (PIDER), que se plasma la noción de desarrollo rural integral, conceptualizada con una visión intersectorial, interinstitucional e instrumentada con un enfoque territorial y de planeación participativa.

Posteriormente, la Ley de Desarrollo Rural Sustentable (LDRS) de 2001 ha modificado sustancialmente la política de desarrollo agropecuario, pesca, alimentación y de desarrollo rural, al reconocer explícitamente que el concepto de lo rural no se reduce al sector agropecuario. La ley expone la importancia del empleo y el ingreso proveniente de las actividades rurales no agropecuarias para la población de este sector. Retoma la sinergia de la coordinación interinstitucional de las dependencias públicas, instrumentada a partir de la creación de programas especiales concurrentes y de una comisión intersecretarial encabezada por la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.

Además, retoma el enfoque territorial del PIDER planteando un desarrollo de la política sectorial de abajo hacia arriba, creando órganos consultivos para la participación ciudadana local, regional, estatal y nacional. Por último, la LDRS refleja un acercamiento integral al sector rural trabajando en cuatro ejes: 1) económico, 2) capital físico –el cual incluye la conservación del medio ambiente–, 3) desarrollo humano –acceso a educación, salud y desarrollo de una cultura del medio ambiente– y 4) capital social. Dado lo anterior, se denota el papel central que tiene y ha tenido la Sagarpa en la conducción de la política de desarrollo rural.

En cuanto al tema de la evaluación de los programas de este sector, es importante mencionar que la Sagarpa es una de las primeras secretarías a nivel federal en someter sus programas a la evaluación externa evaluando un par de éstos en 1998 (ocho años antes de la creación del Coneval). Ambas corresponden a evaluaciones de programas orientados al desarrollo rural y fueron las realizadas al Programa de Capacitación y Extensión Rural y al Programa de Equipamiento Rural. Las evaluaciones corrieron a cargo del Instituto Nacional para el Desarrollo de Capacidades del Sector Rural (Inca Rural) y el Colegio de Postgraduados (Colpos) respectivamente (FAO, 2006).

Es a finales de esta década cuando también se realizan las primeras evaluaciones al programa Alianza para el Campo² por la Organización para la Agricultura y la Alimentación (FAO por sus siglas en inglés) y la Comisión Económica para América Latina. Sin embargo, estas evaluaciones fueron realizadas con metodologías de evaluación distintas entre sí y focalizadas en programas con intereses específicos de la política sectorial.

Por ello, en 1999 al observarse que “se obtenían productos de calidad heterogénea [...] que en ese entonces no existían en la Secretaría de Agricultura, Ganadería y Desarrollo Rural (Sagar)³ antecedentes de procesos sistematizados de evaluación de programas, ni existían capacidades y estructuras desarrolladas a nivel institucional para tal efecto” (FAO, 2006:11) y que se requería evaluar a todos los programas y no sólo algunos de la estrategia “Alianza para el campo”, se estableció un acuerdo de cooperación técnica entre la Sagarpa y la FAO para que esta última realizara las evaluaciones y ofreciera asistencia técnica en el desarrollo de capacidades en los estados.

Derivado de dicho acuerdo, se crea la Unidad de Apoyo FAO (UA-FAO) en 1999, que opera hasta la fecha con consultores seleccionados por FAO México y contratados por este último con recursos derivados de los convenios pactados con la Sagarpa. De este modo, la UA-FAO fue la encargada de realizar evaluaciones a los programas de la Sagarpa del ejercicio 1998-2000, las cuales se dirigían “a vigilar el cumplimiento normativo de metas físicas y financieras” (FAO, 2006:5). Las negociaciones y acuerdos sobre los términos de referencia (TR) de las evaluaciones entre FAO (como evaluador externo) y la Sagarpa (como contratante) se realizaban anualmente. Estos primeros años en la evaluación de los programas de la Sagarpa fueron importantes ya que constituyeron los

² “La Alianza para el campo surgió en 1996 como un conjunto de programas y acciones que engloba las grandes líneas de la estrategia federal de desarrollo del sector agropecuario mexicano, programa descentralizado de subsidios ejecutado por la Sagarpa a nivel federal y a través de sus delegaciones en los estados” (FAO, 2006:8).

³ Por modificaciones al artículo 35 de la Ley Orgánica de la Administración Pública Federal (1994), la Secretaría de Agricultura y Recursos Hidráulicos (SARH) modifica su nombre al de Secretaría de Agricultura, Ganadería y Desarrollo Rural (Sagar) en 1994; y en el 2000 toma el nombre de Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Sagarpa).

primeros ejercicios de aprendizaje y desarrollo de metodologías y TR tanto de las evaluaciones estatales como de las nacionales.

Asimismo, el área de evaluación y análisis de políticas de este organismo internacional se encargó de efectuar las evaluaciones comprendidas en el periodo 2001-2006. Sin embargo, éstas tuvieron cambios considerables respecto a sus objetivos y metodologías dado que, a inicios de este periodo (2001), el Decreto del Presupuesto de Egresos de la Federación (PEF) estableció la obligación de evaluar los programas federales sujetos a reglas de operación; pero “no se preveían elementos, recursos, instrumentos e instancias para su evaluación” (FAO, 2006:11). Por lo anterior, las evaluaciones de este periodo fueron hechas con metodologías propuestas por FAO y realizando una evaluación por cada conjunto de programas. Es decir, las evaluaciones se llevaron a cabo a los grupos de programas de Fomento Agrícola, Ganadero y de Desarrollo Rural; aunque otros programas por sus características particulares fueron considerados individualmente dado que cada conjunto de estrategias era integrado por programas base y programas específicos.

Las evaluaciones mencionadas tuvieron énfasis en el análisis de los procesos y de la gestión, así como de los impactos técnico-económicos y sociales, los resultados, efectos y alcances de la inversión del programa y en los objetivos y metas planteados. Durante este periodo, la evaluación de los programas era vista como una práctica instrumental cuyo objetivo era mejorar la planeación, operación y asignación de recursos para los programas y conocer si se alcanzaban las metas establecidas.

Respecto de sus informes, éstos presentaban la siguiente estructura: una carátula seguida de una breve presentación, una introducción, una descripción del contexto y características del programa, la evaluación de la operación y los resultados-impactos del programa y, por último, las conclusiones, bibliografía y anexos. Las evaluaciones realizadas en este periodo se apoyaban del análisis documental con entrevistas aplicadas a actores clave en los tres órdenes de gobierno y de la sociedad rural, en una muestra de estados seleccionados. Por último, las conclusiones obtenidas eran resultado de un análisis integral del programa de desarrollo rural como un conglomerado de subprogramas en el que se hacía un análisis retrospectivo, un análisis del desempeño del programa en el año de evaluación y un análisis prospectivo del mismo.

Para el caso particular de los programas de fomento al desarrollo rural los programas base para cada año de la estrategia de la alianza para el campo fueron los siguientes:

- En 2001: apoyo al desarrollo rural, desarrollo productivo sostenible en zonas rurales marginadas, mujeres en el desarrollo rural, programa de extensión y servicios profesionales, programa de impulso a la producción de café, fomento a empresas comercializadoras del sector agropecuario y el programa de asistencia técnica al microfinanciamiento rural.
- En 2002-2006: el Programa de Apoyo a los Proyectos de Inversión Rural (PAPIR), el Programa de Desarrollo de Capacidades en el Medio Rural (Prodesca) y el Programa de Fortalecimiento de Empresas y Organización Rural (Profemor).

En el marco del convenio Sagarpa-FAO, se desarrolló un conjunto de metodologías para la evaluación de los programas del sector rural, la sistematización e institucionalización (al menos al interior de la Secretaría) de las evaluaciones externas de los programas de la Sagarpa y el desarrollo de capacidades institucionales de evaluación a nivel estatal.

En el Cuadro 18 se puede observar el número total de evaluaciones realizadas a los programas federales de Sagarpa desde el 2001; en el caso específico del Programa Desarrollo Rural, de 2002-2006, sólo se hizo una evaluación anual, como un conjunto de subprogramas. Como se puede ver, el número de evaluaciones ha aumentado para algunos años desde la creación del Coneval, concentrándose en valoraciones más sintéticas y por lo tanto también más sencillas como lo son las evaluaciones específicas y específicas de desempeño. Además, los procedimientos de contratación, el establecimiento de TR y metodologías comunes para la realización de las evaluaciones, el proceso y el sistema mismo de evaluación actual han cambiado con el establecimiento de los nuevos Lineamientos de 2007.

Otro cambio relevante a partir de 2007 fue que la Sagarpa empezó a contratar a otras entidades como evaluadores externos como lo son consultores privados, otros organismos internacionales, universidades y centros de investigación nacionales. Además, la presentación de los informes también ha cambiado dado

que el Coneval establece el formato, los criterios generales y, en algunos casos, incluso las preguntas de evaluación.

Si bien el Coneval ha contribuido a la estandarización y sistematización de las evaluaciones de los programas federales en México, autores como Fox y Haight realizan una crítica al uso y los impactos que han tenido las evaluaciones externas de la Sagarpa mencionando que éstas “han tenido un impacto limitado en el proceso de formulación de políticas [...] dichas evaluaciones se han centrado en cumplir con las reglas administrativas, sin cuestionar los parámetros que establece la dependencia que se está evaluando” (Fox y Haight, 2010:52-53).

Por otro lado, existen algunos ejemplos de “metaevaluaciones” realizadas de manera independiente a las evaluaciones de los programas federales de la Sagarpa. Sin embargo, estos trabajos no necesariamente constituyen una verdadera⁴ metaevaluación dado que no analizan evaluaciones; más bien constituyen análisis que intentan valorar el impacto de los programas. Los estudios a los cuales se hace referencia son los elaborados por Vargas (2007) y Schwentesius *et al.* (2007). El primero es un documento solicitado y financiado por el CEDRSSA de la Cámara de Diputados; el cual se ubica a sí mismo en el marco del NPM como “una reflexión sobre el enfoque y métodos empleados, de sus resultados y sus recomendaciones” (Vargas, 2007:13).

El segundo es un estudio comparativo del Programa de Pago Directo Procampo en tres años (1998, 2001 y 2003). En el ejercicio los autores recuperan definiciones de metaevaluación en las cuales resaltan que la misma tiene como objetivo analizar la calidad de la evaluación y el desempeño de quienes la realizan. Sin embargo, en otro apartado de dicho estudio se expone que el objetivo es “identificar qué diferencia ha marcado el programa en comparación con los no beneficiarios y qué impacto ha tenido sobre los beneficiarios” (Schwentesius *et al.*, 2007:110) lo que correspondería más bien a un diseño experimental de evaluación de impacto.

⁴ No constituye una metaevaluación al retomar la definición propuesta por Cardozo (2012a) en la que se define como un tipo particular de evaluación, el cual enjuicia el valor y el mérito no sólo de los trabajos de los evaluadores, sino el de todo el sistema en que se organiza y desarrolla el anterior, incluyendo el papel que tienen otros actores o instancias que inciden en la calidad de todo el proceso desarrollado y del trabajo final logrado.

CUADRO 18
Evaluaciones de programas de Sagarpa 2001-2013

Tipo de evaluación	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Complementaria								1	1					2
Consistencia y resultados							11				15			26
Diseño							1	6		1				8
Específica												10	15	25
Específica de desempeño								9	8	8		10		35
Estratégica									1					1
Otras	7	9	10	13	11	8								58
Total	7	9	10	13	11	8	12	16	10	9	15	20	15	155

Fuente: elaboración propia con datos del Coneval [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeResultados.aspx>] [<http://www.fao-evaluacion.org.mx/pagina/informe-evaluacion-programas>].

Por último, en el trabajo de Medina (2014) se realizan tres metaevaluaciones, una por cada una de las ECR 2011-2012 hechas a los programas de: 1) Apoyo al Ingreso Agropecuario: Procampo para Vivir Mejor; 2) Prevención y Manejo de Riesgos y 3) Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural; evaluaciones realizadas por la FAO, El Colegio de México (Colmex) y el Centro de Investigaciones y Estudios Superiores en Antropología Social CIESAS-Occidente, respectivamente.

Respecto al enfoque metodológico utilizado se retoma el propuesto por Daniel Stufflebeam (1974, 1999, 2001), en el cual, mediante la contrastación de la información dispuesta en las evaluaciones, la información documental complementaria y la información cualitativa recabada (plasmada en una lista de chequeo) ante un conjunto de estándares,⁵ se emite una valoración de las mismas en términos de la utilidad, factibilidad, legitimidad y precisión de cada evaluación. Aunado a lo anterior, se otorga una calificación cuantitativa a la evaluación que refleja la calidad⁶ de la misma.

En conclusión, la evaluación del desarrollo rural ha sido de las primeras aplicadas (1998) y contó con un esquema especial de cercanía entre evaluador y evaluado durante 2001-2006 (convenio con FAO) que se modificó después de la creación del Coneval, dando lugar a la contratación de nuevos equipos evaluadores que realizaron un promedio de 14 evaluaciones externas anuales.

⁵ Los estándares que guían las metaevaluaciones son: utilidad, factibilidad, legitimidad y precisión de la evaluación y son aquellos propuestos por el Joint Committee on Standard for Educational Evaluation (JCSEE). Los estándares se pueden aplicar a todas las evaluaciones, pero las formas exactas que se aplican serán diferentes, ya que dichas normas no están diseñadas para ser aplicadas igualmente a todos los programas.

⁶ La calidad de la evaluación es medida por una asignación cuantitativa otorgada según el cumplimiento de los estándares provistos en cada uno de los cuatro criterios propuestos por el Joint Committee on Standard for Educational Evaluation.

Conclusiones generales de la primera parte

Myriam Cardozo Brum

Con la intención de simplificar la integración de la información generada a lo largo de los capítulos que componen esta parte y luego analizarla, se elaboró el Cuadro 19, donde se puede ver que, entre los países pioneros de Europa y Norteamérica, destaca Estados Unidos de América porque su proceso inició aproximadamente seis o siete décadas antes que los demás; cuenta con un alto grado de institucionalización de la evaluación, con cinco revistas especializadas y con la mayor oferta de formación universitaria.

En materia de institucionalización le sigue Reino Unido, cuyo trabajo inicia en la década de 1970 y, al igual que en Estados Unidos de América, su impulso proviene tanto del Poder Ejecutivo como del Legislativo. En Canadá la institucionalización, iniciada casi al mismo tiempo que en Reino Unido, es fuerte en la administración pública y la única con procedimientos para garantizar su calidad, pero el Parlamento se involucra poco. Francia inició la incorporación de la evaluación una década antes, con esfuerzos alternados de su Poder Ejecutivo y su Asamblea Nacional, y ha modificado sus enfoques en un proceso de búsqueda de uno más cercano a su cultura que el utilizado en los países antes referidos, fuertemente formalizado y experimental. Sin embargo, recientemente parece reorientarse nuevamente hacia ellos.

En los países latinoamericanos, los procesos de evaluación inician desde la década de 1970 en Costa Rica, pero hoy se puede considerar que Colombia, Chile y México han avanzado más en su institucionalización, a pesar de haber comenzado una o dos décadas después. En Colombia el trabajo se encuentra

muy consolidado, el de México es muy similar pero bastante posterior y el de Chile está avanzado rápidamente y con mayor transparencia (i.e. concursos públicos). El enfoque experimental y econométrico tiene mucho arraigo en estos tres países, aunque México reconoce las dificultades de su aplicación, por lo que ha realizado pocas evaluaciones de impacto. El proceso de Costa Rica ha sido lento, tiene pocas evaluaciones y fundamentalmente son internas. En general, la calidad de los trabajos ha sido poco cuidada.

En materia de asociaciones de profesionales de la evaluación, todos los países del primer grupo han formado una con actividades diversas: conferencias anuales, grupos de discusión, secciones regionales, etcétera. Francia destaca por la intensidad de su trabajo reflexivo y por tener, al igual que Canadá, sociedades regionales importantes, además de la nacional (i.e. Lille en Francia, Quebec en Canadá). En los países latinoamericanos, salvo la Asociación Centroamericana de Evaluación, las asociaciones de profesionales no existen. Sólo se dispone de redes nacionales (Redcolme, Redmexme) que tanto difunden información, publicaciones y eventos como organizan algunos grupos de discusión a través de internet, pero únicamente ofrecen oportunidades de intercambios cara a cara cuando la Relac o la Redlacme organizan eventos internacionales periódicos (que van de uno a cinco años).

La situación es muy desigual en cuanto a publicaciones especializadas: mientras Estados Unidos de América tiene cinco revistas, Francia no tiene ninguna; en cambio en materia de libros los dos países cuentan con mayor cantidad de publicaciones que Reino Unido y Canadá. Por su parte, ningún país latinoamericano dispone de una revista en temas de evaluación en general (sólo aplicados a casos como los educativos) y los libros publicados son pocos, resaltando México con unos 15, algunos de éstos relativos a la metaevaluación, tema también tratado en Costa Rica.

En cuanto a formación universitaria en evaluación, en el primer grupo vuelve a destacarse Estados Unidos de América, en este caso junto con Canadá, mientras Francia sólo tiene una maestría especializada en evaluación en general. En el segundo grupo, Chile y Costa Rica también ofrecen una maestría, mientras México acaba de iniciar su primera especialización de un año de duración. Todos ofrecen adicionalmente diplomados y cursos esporádicos.

En síntesis, Estados Unidos de América y Reino Unido tienen un sistema consolidado de evaluación con participación de los Poderes Ejecutivo y Legislativo,

CUADRO 19

Síntesis de características de la evaluación por países

Institucionalización	Asociaciones profesionales	Publicaciones especializadas	Programas académicos	Fortalezas
ESTADOS UNIDOS DE AMÉRICA				
Pais pionero con inicios a principios de siglo XX. Cultura nacional de evaluación en todos los niveles, alta formalización, abundantes recursos, calidad parcialmente supervisada.	American Evaluation Society. Realiza una conferencia anual y tiene 51 grupos de discusión.	Cinco revistas (<i>American Journal of Evaluation, Evaluation Review, New Directions for Evaluation, Evaluation and Program Planning and Journal of Multidisciplinary Evaluation</i>) y numerosos libros.	Amplia oferta de cursos, maestrías y doctorados específicos en evaluación.	Consolidación en los poderes Ejecutivo y Legislativo mediante órganos especializados (CBO, GAO, OMB).
REINO UNIDO				
Inició en la década de 1970. Institucionalización amplia, con participación del Poder Ejecutivo y el Parlamento, ligada a la eficiencia, con calidad parcialmente supervisada.	United Kingdom Evaluation Society (UKES). Realiza una conferencia anual y está ligada a la Sociedad Europea de Evaluación.	Revista <i>The Evaluator</i> de la UKES, algunas guías y pocos libros.	Maestrías en políticas públicas con cursos sobre evaluación.	Avances importantes en todos los Departamentos del gobierno y órgano especializado en evaluación en el Parlamento (NAO).
FRANCIA				
Inicio en la década de 1960. Esfuerzos alternativos de los Poderes Ejecutivo y la Asamblea Nacional con cambios de enfoque. Calidad asegurada mientras existió el CSE (1990-1998).	Société Française d'Évaluation. Organiza continuas jornadas de estudio y 4-5 eventos relevantes por año. Cuenta con un Observatorio, redes, grupos y diversos clubes regionales.	Abundantes libros y varios artículos, pero no cuenta con revistas especializadas.	Maestrías y doctorados con cursos en evaluación (políticas públicas, administración, economía) y sólo una maestría especializada.	Amplia reflexión y actividades en las SFE. Orientada a la investigación y la evaluación externa con énfasis en el aprendizaje de actores.

continúa

CUADRO 19

Síntesis de características de la evaluación por países

Institucionalización	Asociaciones profesionales	Publicaciones especializadas	Programas académicos	Fortalezas
	CANADÁ			
<p>Inició en la década de 1960. Cultura de la evaluación interna entre funcionarios para la modernización de la gestión. Sistema no obligatorio. Calidad asegurada por Tesoro y Corte de Cuentas.</p>	<p>Canadian Evaluation Society (CES) con conferencia anual y 11 secciones regionales (2003), especialmente la Société Québécoise d'Évaluation de Programme (SQEP).</p>	<p><i>Canadian Journal of Program Evaluation</i> y <i>Revue Internationale en Évaluation du Développement</i> más algunos artículos en otras revistas y pocos libros.</p>	<p>Muy abundante y especializada a nivel nacional. Cursos en programas más amplios en Quebec (políticas públicas, asuntos públicos, administración, medio ambiente).</p>	<p>Fuerte sistema intragubernamental. Promoción por el Tesoro y la Corte de Cuentas con poca participación del Parlamento.</p>
	COLOMBIA			
<p>Avance importante de la cultura orientada a resultados y rendición de cuentas.</p>	<p>Red Colombiana de Monitoreo y Evaluación (Redcolme), asociada a Relac y Redlacme. Encuentro nacional en 2013 y Conferencia Internacional en 2007.</p>	<p>No dispone de revistas especializadas. Pocos artículos y libros.</p>	<p>Cursos de evaluación en programas universitarios más amplios (ingeniería, economía, gestión, ciencias políticas). Pocas jornadas de capacitación intragubernamental.</p>	<p>Inició en la década de 1980. Sistema consolidado de evaluación externa, con poca utilización de sus resultados y garantía de libre competencia de los evaluadores.</p>
	COSTA RICA			
<p>Cultura parcialmente institucionalizada de evaluación interna, con pocas evaluaciones externas.</p>	<p>Asociación Centroamericana de Evaluación con especial presencia de Costa Rica. Organizó la Conferencia de Relac 2010 y de Redlacme 2012.</p>	<p>No dispone de revistas especializadas. Pocos artículos y libros. Avances en metaevaluación.</p>	<p>Maestría en evaluación de programas y proyectos de desarrollo y cursos en <i>master</i> profesional en población y salud. Fuerte capacitación impulsada por el gobierno.</p>	<p>Inició en la década de 1970. Proceso lento pero fuerte en evaluación interna y utilización de sus resultados.</p>

continúa

CUADRO 19

Síntesis de característica de la evaluación por países

Institucionalización	Asociaciones profesionales	Publicaciones especializadas	Programas académicos	Fortalezas
CHILE				
Sistema muy organizado de evaluación y control de gestión. Evaluación externa por universidades y consultoras mediante concursos públicos. Énfasis en resultados e impactos (medición experimental).	Sólo Asociación Chilena de Políticas Públicas. Algunos de sus evaluadores tienen fuerte presencia en la Relac.	Artículos en revistas no especializadas. No se identificaron libros sobre el tema, a lo sumo capítulos en publicaciones sobre temas más amplios como políticas públicas.	Dos maestrías especializadas, otra de medio ambiente con algunos contenidos en evaluación, cursos y diplomados.	Inicia en la década de 1990 pero avanza aceleradamente. Ampla difusión y utilización promovida con incentivos. Compromisos y seguimiento institucional de acuerdos en torno a resultados.
MÉXICO				
Creación de órganos especializados a nivel federal (Coneval, INEF, Coordinación dentro del INECC) y local (Evalúa DF y otros). Existe incertidumbre a raíz de recientes cambios en la normatividad del Coneval. Evaluaciones externas con pocos concursos y limitado control de calidad.	Cuenta con un capítulo mexicano de la Redlame (Redmexme), la Relac tiene poca presencia. Existe la Asociación Mexicana de Profesionales de la Evaluación pero no representa a todos los sectores ni tiene cobertura nacional.	Se publican artículos en revistas diversas, no existe una especializada. Se han publicado unos 15 libros sobre el tema.	Acaba de iniciarse una especialidad en evaluación integral de programas y políticas de desarrollo social. Se cuenta con algunas maestrías en administración y políticas públicas con contenidos de evaluación. Además existe una oferta limitada de diplomados y cursos.	Inició finalizando el siglo pasado y se fue ampliando aceleradamente en 2001 y 2007. Existen pocos análisis de la calidad de los trabajos y limitada utilización de resultados.

Fuente: elaboración propia con base en los capítulos previos.

mientras el canadiense es importante pero sólo intragubernamentalmente. En cambio Francia destaca por sus esfuerzos en el campo de la investigación, la reflexión y la discusión profunda del tema y por haber desarrollado un enfoque alternativo al énfasis econométrico y experimental: constructivista con mayor peso de la explicación cualitativa y de la participación y aprendizaje de todos los actores.

El inicio más tardío de los procesos latinoamericanos de evaluación y su mayor restricción de recursos (financieros y humanos capacitados) explican sus avances más limitados. Si bien Costa Rica fue el primero, casi al mismo tiempo que Francia y Canadá, y sigue siendo pionero en formación universitaria y espacios asociativos de profesionales de la evaluación, su proceso a nivel gubernamental ha sido lento, lo que ha provocado que los otros países analizados alcancen actualmente una mayor consolidación. Éstos han institucionalizado sistemas de evaluaciones externas muy similares: Colombia es el más antiguo, Chile el más acelerado y en el caso de México el proceso será revisado a continuación, por sectores (Cuadro 20).

CUADRO 20
Síntesis de las características de la evaluación por sectores mexicanos

Sector desarrollo social	Sector educación	Sector medio ambiente	Sector salud	Sector desarrollo rural
Intensa experiencia pero aún con muchos retos pendientes.	Concentración del mayor número de evaluaciones, no está clara la delimitación de las atribuciones entre el INEE y el Coneval.	Con avances, pero con temas poco cuidados (ANP, cambio climático) y una coordinación de evaluación recientemente creada.	Con experiencia en monitoreo de indicadores, un número importante de evaluaciones y sin órgano especializado que las organice.	Pionero en el inicio de las evaluaciones, sistema convenido con FAO hasta 2006. Tampoco cuenta con un órgano especializado que coordine sus evaluaciones.

Fuente: elaboración propia con base en los capítulos previos.

En México, las secretarías con más experiencia en la realización de evaluaciones han sido la Sedesol y la Sagarpa, si bien la SEP duplica el número de las realizadas en 2007-2013. En el primer caso fue importante el esfuerzo realizado entre 2001 y 2006 en que se intentaron trabajos de evaluación integrales, con equipos de reciente formación puestos a competir mediante invitaciones a universidades y centros de investigación. En el segundo, las evaluaciones fueron realizadas por convenio con la FAO. Educación, Salud y Medio Ambiente, aunque presentaban algunos casos, se integraron plenamente al proceso de evaluación a partir de la creación del Coneval en 2006, resaltando educación por su elevado número (37% del total en 2012). Este nuevo periodo se ha caracterizado por un notable incremento de la asignación directa de contratos (sin concurso) a las instituciones ya mencionadas y también a las consultorías.

Como se mencionó, el número de programas federales cuya evaluación coordina el Coneval se incrementó en forma acelerada y también ha empezado a brindar asesoría a los gobiernos locales, incluso a evaluar sus avances. Este crecimiento de sus tareas ha ido acompañado de una enorme simplificación de los trabajos realizados y hace cuestionarse sobre la capacidad de dicho Consejo para no caer en privilegiar la cantidad en detrimento de la calidad. Esto podría resolverse en parte mediante la creación de órganos especializados en evaluación como el INEE en educación o, al menos, una Coordinación como en el caso del INECC en medio ambiente, cuidando de delimitar bien sus atribuciones. Seguramente salud ameritaría un esfuerzo similar y también desarrollo social cuya experiencia acumulada es importante y constituyó la base de despegue del Coneval. Eso permitiría poner mayor atención al conjunto de programas relevantes de cada sector, lo que se detecta necesario en el caso de la Secretaría de Medio Ambiente y Recursos Naturales.

Para finalizar, es importante subrayar que de los países revisados, sólo Canadá (Francia lo hizo temporalmente; México únicamente a nivel local en 2008-2009) se ha preocupado de analizar sistemáticamente la calidad de sus trabajos evaluativos, lo cual resulta relevante para decidir sobre la aplicación de sus recomendaciones. La segunda parte de este libro expondrá metaevaluaciones dirigidas a valorar la calidad de los trabajos mexicanos y la tercera parte analizará la utilización de sus resultados y sus posibles explicaciones, incluida la calidad.

SEGUNDA PARTE
Metaevaluación: calidad de las evaluaciones
aplicadas en México 2007-2014

Introducción

La primera parte de este libro permitió conocer los avances alcanzados por un conjunto de países seleccionados en materia de institucionalización y profesionalización de sus procesos de evaluación y sus principales resultados. En su presentación fueron utilizados indicadores cuantitativos como el número de evaluaciones realizadas, pero no se profundizó en el análisis de la calidad de las mismas que, como ya se mencionó, resulta relevante para decidir sobre la aplicación de sus recomendaciones.

El estudio de dicha calidad en el caso mexicano constituye el objetivo de esta segunda parte que inicia con una descripción de la normatividad que guía los procesos de evaluación a nivel nacional y del Gobierno del Distrito Federal, para continuar con el diseño de un nuevo ejercicio de metaevaluación y extenderse en la presentación de los resultados.

Dado el tamaño del universo de evaluaciones realizadas en el periodo 2007-2014 fue imposible trabajar con una muestra representativa de las mismas por lo que, como se justifica en el apartado correspondiente, se incluye el análisis de un total de 16 evaluaciones a nivel federal que fueron seleccionadas de manera de cubrir las áreas de desarrollo social, educación, medio ambiente, salud y desarrollo social (una realizada directamente por la instancia responsable del programa y el resto incluido en el programa anual del Coneval), los diferentes tipos de evaluación (diagnóstico, diseño, procesos, impactos, etcétera) y su distribución anual. Además se incluyeron cinco evaluaciones realizadas a nivel

del Gobierno del Distrito Federal en materia de política social, transporte y medio ambiente (también una contratada directamente y las otras coordinadas por Evalúa DF).

El trabajo se realizó fundamentalmente mediante el análisis documental de términos de referencia e informes de evaluación y se complementó con la aplicación de algunas entrevistas semiestructuradas o solicitudes de información a informantes-clave.

CAPÍTULO XIV

Evaluación y metaevaluación

Myriam Cardozo Brum

En trabajos previos se ha definido a la evaluación como un proceso consistente 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”. También se ha considerado a la metaevaluación como un tipo particular de evaluación que “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” (Cardozo, 2012a:15, 17).

La diferencia principal radica en que mientras la evaluación toma por objeto de estudio a las políticas y programas públicos, la metaevaluación se concentra en analizar la calidad de las evaluaciones (independientemente del desempeño de dichas políticas y programas). Es decir que una evaluación puede ser calificada por su alta calidad aunque haya arrojado un mal desempeño del programa y viceversa.

Para proceder a la metaevaluación de las políticas y programas incluidos en esta parte se comienza con una referencia mínima a la normatividad general vigente a nivel federal y del GDF como parte fundamental del contexto en que se desarrollan los procesos estudiados.

Normatividad y tipos de evaluación

En materia de evaluación a nivel federal, la LGDS publicada el 20 de enero de 2004 prevé su realización 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 (artículo 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 (artículos 74-76). La ley enfatiza la importancia de la participación social y ciudadana mediante el mecanismo de contralorías sociales (artículos 61-71) y determina que la evaluación será anual y también podrá ser multianual (artículo 78).

El Reglamento de la ley anterior establece en su artículo 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 éste establezca (artículo 77). Por su parte, el artículo 79 aclara que las evaluaciones de impacto de los programas de desarrollo social podrán realizarse mediante 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 evaluación anual.

En síntesis, la LGDS y su reglamento exigen evaluaciones anuales que midan: 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 periodo 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 2006 fue creado el Coneval en tanto organismo público descentralizado, sectorialmente coordinado por la Sedesol, encargado de revisar periódicamente

el cumplimiento de objetivos de los programas y de las acciones de la política de desarrollo social. Como se mencionó en el capítulo 8, el 30 de marzo de 2007 este Consejo publicó en el DOF, conjuntamente con la SHCP y la SFP, los Lineamientos que sustituirán el Acuerdo de 2002 que normaba la realización de las evaluaciones y había continuado vigente hasta ese momento. Su objetivo es regular: la evaluación de los programas federales de todos los sectores vinculados a la política de desarrollo social nacional (artículo 72 de la LGDS), 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.

En lo que se refiere concretamente a las evaluaciones, éstas son 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 (Cuadro 21):

- ECR, que analiza el diseño, desempeño global y resultados basados en el seguimiento de indicadores, mediante trabajo de gabinete. Cuenta con TR genéricos (Coneval, 2007b; 2011h).
- Evaluación de indicadores, que estudia la pertinencia y el alcance de los indicadores para la medición de resultados, incluyendo trabajo de campo.
- Evaluación de procesos, que revisa la eficacia y eficiencia de los procesos operativos y su contribución al mejoramiento de la gestión; cuenta con TR genéricos (Coneval, 2009a; 2013i).
- Evaluación de impacto, que mide 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 (Coneval, s/f:1).
- Evaluación específica, que incluye todas las no comprendidas en los tipos previos, que pueden implicar tanto trabajo de gabinete como de campo

Se agrega que los principales contenidos de las ECR se ajustarán a TR genéricos e incluirán (capítulo II, artículo décimo octavo):

- *Diseño del programa*: implica la identificación de un problema o necesidad pública no satisfecha que el programa puede resolver; la contribución del mismo al logro de los objetivos estratégicos incluidos en planes y programas; la evidencia científica de que los productos o servicios del programa contribuyen al logro de su fin; la lógica vertical y horizontal de los componentes del marco lógico (objetivos, estrategias, indicadores de seguimiento, etcétera); la definición de la población potencial y objetivo así como su justificación; la estructura del padrón de beneficiarios; las RO aplicadas; y la posible complementariedad o duplicidad con otros programas federales.
- *Planeación estratégica*: contiene mecanismos y herramientas de planeación, de establecimiento y definición de metas e indicadores, la posibilidad de medir rigurosamente su impacto, la difusión y utilización de los resultados y evaluaciones.
- *Cobertura y focalización*: conlleva la cuantificación de la población potencial y objetivo; los métodos empleados para lograrla; la población atendida en relación con la potencial y objetivo; la efectividad de la estrategia de cobertura y focalización; y sus mecanismos.
- *Operación*: requiere la revisión del cumplimiento de las RO; acciones de mejora y simplificación regulatoria; mecanismos de organización y gestión; administración financiera de recursos; eficacia, eficiencia y economía del programa; sistematización de información; mecanismos de actualización y depuración del padrón de beneficiarios; cumplimiento y avance de indicadores; difusión y rendición de cuentas.
- *Percepción de la población objetivo*: se compone del análisis del nivel de satisfacción de la población objetivo y otros actores clave con base en sus percepciones.
- *Resultados*: incluye la medición del grado de alcance de los propósitos y fines del programa; la calidad de las evaluaciones de impacto aplicadas; el rigor con que puede realizarse la anterior; la difusión de los resultados y evaluaciones del programa; y la utilización de la información generada para mejorar su desempeño.

Las evaluaciones de diagnóstico y de diseño (independientes de las ECR que incluyen este último contenido) de programas nuevos están previstas en el capítulo IV, artículo vigésimo primero y vigésimo segundo de los Lineamientos, mientras las complementarias aparecen en el capítulo V, artículo vigésimo cuarto.

Los elementos mínimos a considerar en la realización de las evaluaciones de diagnóstico fueron establecidos hace ya varios años (Coneval, 2011a), pero recién en 2014 la Sedesol 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. Están referidas a programas nuevos como los comedores comunitarios pero, aunque se suponen previas a la implementación, también se han aplicado a otros de larga data como coinversión social.

Las evaluaciones de diseño cuentan con TR genéricos (Coneval 2007a y s/f:2); retoman un instrumento de planeación del marco lógico utilizado por varios organismos internacionales, que en México fue transformado en la MIR. Según el Programa para un Gobierno Cercano y Moderno 2013-2018, su utilización ha permitido cubrir en forma creciente el gasto programable: 49.14% promedio en el periodo 2008-2012 y 69.5% en el último de dichos años. Sin embargo, también afirma que “las MIR no se encuentran totalmente vinculadas con el proceso de planeación nacional, lo que provoca una generación excesiva de indicadores de desempeño, a su vez se encuentran desvinculados con los programas sectoriales”, por lo que se ha procedido a alinearlas con base en el Plan Nacional de Desarrollo (PND) actualmente vigente.

Las evaluaciones complementarias de desempeño son consideradas en los Lineamientos como 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 términos de referencia.

Por otra parte, no se ha realizado ninguna evaluación identificada como de indicadores ni existe una metodología explícita para realizarlas, y a la fecha sólo 27 se han enfocado a procesos;¹ sin embargo, las primeras pueden formar parte del análisis de las MIR (salvo que ésta no considera la realización de trabajo de campo que este tipo de evaluación preveía); mientras las segundas pueden quedar incluidas cuando se analiza la operación en las ECR y, en el caso particular de

¹ Según la SHCP eran sólo 14 en 2011 y Székely (2013) reporta 16 en 2012.

Oportunidades, existe una evaluación de la operación que aparece como parte de la evaluación de impacto 2008 (tomo IV).

Además en 2008, dentro de las evaluaciones específicas, el Coneval comenzó a contratar las denominadas Específicas de Desempeño, para las que recién se elaboró un modelo de TR en la generación 2012-2013, que solicita: datos generales, resultados/productos, evolución de la cobertura, seguimiento a ASM, y datos de la instancia evaluadora. Se trata prácticamente de un modelo de monitoreo de algunos indicadores, más que de una evaluación. En él se expresa que “se debe realizar mediante trabajo de gabinete y únicamente con base en la información proporcionada por el Coneval” (Coneval, 2013a:7). Normalmente se indica cuál es el mínimo de información exigido, pero en este caso prácticamente se prohíbe usar información de fuentes distintas al propio Coneval, lo cual se considera una restricción grave a la libertad del evaluador.

A este conjunto de evaluaciones, en 2009 el Coneval agregó las integrales de desempeño, que si bien 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 (Coneval, s/f:4). Székely (2013) reporta que se realizaron 19 evaluaciones integrales en 2009 y 21 en 2010, las que no fueron reportadas por la SHCP (Cuadro 11), no están incluidas en el PAE ni en los Lineamientos. En síntesis, puede haber evaluaciones de desempeño específicas y complementarias de programas o las que integran un conjunto homogéneo de ellos.

Si se contrasta la tipología revisada con los contenidos exigidos por la LGDS y su reglamento, es posible sostener que el tipo de evaluación más completo corresponde a la ECR, por lo que debería constituir la evaluación anual fundamental; sin embargo, sólo se han aplicado en 2007 y 2011. Este tipo incluso agrega la necesaria evaluación de diseño y de la planeación estratégica. La realización de evaluaciones complementarias, estratégicas, específicas, de desempeño o de diseño puede ser útil pero su aplicación no cubre todos los aspectos exigidos anualmente por la Ley General de Desarrollo Social.

Para efectuar una evaluación integral se tendría que cubrir los antecedentes del problema y política o programa evaluado, el enfoque teórico que lo avala dentro del conjunto de los posibles, la situación económica, social y política en

CUADRO 21
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 (2009a, 2011a y 2013a).

el marco de la cual se evalúa, la organización de la evaluación, la formulación (diseño) e implementación (operación) de la política o programa, y sus resultados e impactos. En consecuencia, la metaevaluación debe revisar si todo esto se ha evaluado y si el trabajo realizado cuenta con la calidad necesaria; para lo que también deberá incluir el análisis de las recomendaciones de la evaluación y su seguimiento. Sin embargo, se debe tomar en cuenta que la normatividad revisada introduce limitaciones:

1. Como ya se mencionó, no todos estos aspectos son recuperados por la LGDS y su reglamento, que se concreta al proceso seguido por la política o programa evaluado: logro de objetivos y metas (eficacia), gestión realizada (procesos), cobertura y calidad de los servicios (resultados) e impactos.
2. Los Lineamientos emitidos en 2007 se proponen explicitar y ampliar los cuatro tipos de evaluación previstos; sin embargo, tampoco incluyen el estudio de todo lo que constituye el contexto de la evaluación realizada y se siguen limitando a la revisión de las etapas básicas del proceso de un programa, estrategia o política, retomando en algunos casos el seguimiento de las recomendaciones o Aspectos Susceptibles de Mejora.

3. Se optó por denominar de diferente manera a la reunión de algunos tipos básicos previstos en la LGDS, produciendo un conjunto que no es netamente diferenciable:

- Los lineamientos conservan con claridad tres tipos básicos en materia de programas (diseño, operación e impacto); pero dejan fuera la evaluación de resultados. También se previó un tipo particular de diseño y rediseño, no del programa sino de su sistema de indicadores; pero como ya se mencionó, este tipo no llegó a aplicarse en el marco del Coneval.
- Las combinaciones retoman diseño, operación y resultados de programas en las llamadas Evaluación de Consistencia y Resultados.
- Algunos tipos se definen por exclusión o complementación de las expresamente establecidas para los programas, por lo que no se especifica qué parte del proceso estarán cubriendo. Es el caso de las evaluaciones específicas y las integrales de desempeño.
- Las que van más allá de los programas, refiriéndose a conjuntos de programas, estrategias, políticas e instituciones: estratégicas y complementarias no se diferencian con claridad: mientras las primeras incluyen las estrategias, políticas e instituciones relacionadas con un programa o conjunto de programas; las segundas 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. De esta forma, en las estratégicas está contenida la evaluación de políticas relacionadas con un conjunto de programas que aporten al mismo tema, lo cual es el objetivo de las complementarias.

En síntesis, las categorías usadas en los Lineamientos producen más confusión que si sólo se explicitaran más claramente las combinaciones de los cuatro tipos básicos previstos en la LGDS, definiendo en cada caso si la unidad de análisis será el programa o se ampliará a políticas, estrategias o instituciones.

Si bien el BM elaboró un documento denominado “Análisis de los Lineamientos generales para la evaluación de los Programas Federales de la Administración Pública Federal” que se puede consultar en la página del Coneval, 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, etcétera.

En la práctica se han realizado regularmente las evaluaciones de diseño de nuevos programas (108 hasta 2011). El resto de las evaluaciones se ha concentrado en las EED que son las más simples (396 realizadas hasta 2011) y en las ECR (246, según Székely, 2013), mientras que las de impacto, las más complejas, alcanzaban apenas a 16 en 2007-2011 (Cuadro 11). Los Lineamientos sólo definen (por descarte) a las evaluaciones específicas, pero no mencionan el caso concreto de las EED, que han venido incluyendo básicamente el avance de los indicadores y del logro de objetivos y metas, la cobertura, los resultados, los efectos y hallazgos relevantes del programa, pero no incluyen el estudio de la gestión.

En 2012-2013 el Coneval informó la realización de 80 evaluaciones, principalmente EED, mientras que en 2013-2014 recurrió a la elaboración de evaluaciones específicas reducidas a Fichas de Monitoreo y Evaluación, un esquema simplificado de monitoreo de información que no implica estrictamente una evaluación, al punto que la Dirección de Oportunidades presentó una posición institucional argumentando en su contra debido al poco espacio que las mismas brindan para valorar un programa tan complejo y la restringida justificación de incluir en él un análisis FODA. Sin embargo, Coneval argumenta que las EED y las mencionadas fichas se diseñaron para tomadores de decisión que necesitan información válida, homogénea y comparable de distintos programas.² El problema es que éstas han tendido a sustituir a las evaluaciones más completas, por lo que se concluye que, con el incremento tan importante del número de evaluaciones desde la creación del Coneval en 2006, la tendencia ha sido a simplificar su tipo (al punto de sustituir algunas por un sistema de monitoreo) y estandarizar su diseño.

Por su parte, el artículo 42 de la LDSDF y el Reglamento correspondiente (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, en 2008 quedó instalado el Consejo

² Respuesta del Coneval al cuestionario diseñado para esta investigación.

de Evaluación del Desarrollo Social (Evalúa DF) como organismo público descentralizado, el que ha organizado un sistema de evaluación sustantivamente diferente al federal (Cardozo, 2011 y 2012b). En contraste del de nivel nacional, su normatividad no lo obliga a evaluar sus programas anualmente y establece que sus recomendaciones tienen carácter vinculatorio.

Para guiar la realización de sus trabajos, Evalúa DF publicó una serie de lineamientos disponibles en su página y un pequeño libro de orientación que se refiere a los tipos de evaluación encarados por este Consejo, tanto de programas como de políticas y servicios. Cada uno de éstos 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). En la práctica, la mayoría ha girado en la evaluación de programas y sólo una ha cubierto el impacto (Cuadro 12).

Finalmente, es importante mencionar un esfuerzo no gubernamental por evaluar el conjunto de programas federales. Se trata del trabajo realizado por Gestión Social y Cooperación, AC (Gesoc), un organismo de la sociedad civil especializado en análisis de políticas públicas y evaluación de proyectos, que calcula cada año un Índice de Desempeño de los Programas Públicos Federales (Indep), que varía entre 0 (opaco o sin información) y 100 (óptimo). El mismo se construye a partir de tres indicadores: calidad del diseño, eficacia en el logro de metas y cobertura alcanzada (Gesoc, 2014). En la práctica, en 2014, los resultados de los programas con información variaron entre 96.11 y 7.17, pero un notorio número de ellos (63 de los 182 evaluados, que reciben 9% del gasto social federal) fue ubicado en la zona opaca por carecer de datos para proceder a su evaluación, especialmente en materia de población cubierta. En el Cuadro 22 se presenta la información de un subconjunto de programas ligados a los sectores en los que se procederá posteriormente a la metaevaluación de sus programas (desarrollo social, educación, medio ambiente, salud y desarrollo rural).

El cuadro citado muestra que los programas más cumplidos provienen especialmente de desarrollo social (donde Prospera, antes Oportunidades, alcanza sólo el quinto lugar) y salud. Educación resalta por sus problemas de cobertura y medio ambiente por la falta de información. La Sagarpa, pionera en la evaluación del desarrollo rural en el país, presenta algunos problemas de diseño y cobertura, pero fundamentalmente de opacidad en 80% de sus programas (aunque el cuadro sólo muestra uno como ejemplo).

CUADRO 22
Índice de desempeño de algunos programas públicos federales 2014

Sector	Programa	Índice	Debilidad principal
Desarrollo social	Pensión Adultos Mayores	96.11	Diseño
Salud	Caravanas de la Salud	93.41	Diseño
Desarrollo social	Abasto Rural	93.03	Cobertura
Salud	Seguro Popular	88.02	Eficacia
Desarrollo social	Prospera	86.33	Cobertura
Desarrollo rural	Procampo para vivir mejor	82.90	Diseño
Educación	Escuela Segura	76.17	Cobertura y eficacia
Salud	Sistema Integral de Calidad en Salud	74.02	Diseño y cobertura
Educación	Escuelas de Calidad	59.27	Cobertura
Desarrollo social	Estancias Infantiles	57.69	Cobertura
Educación	Escuelas de Tiempo Completo	52.39	Cobertura
Educación	Escuelas Dignas	44.58	Diseño, cobertura y eficacia
Desarrollo social	Subsidios Programas para Jóvenes	7.17	Diseño, cobertura y eficacia
Medio ambiente	Vigilancia Comunitaria de ANP	0	Opaco
Medio ambiente	Monitoreo Biológico de ANP	0	Opaco
Desarrollo rural	Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo	0	Opaco
Desarrollo social	Fomento y Desarrollo Culturas Indígenas	0	Opaco
Medio ambiente	Mitigación y Adaptación al Cambio Climático	0	Opaco

Fuente: elaboración propia con datos de Gesoc (2014).

Esta forma alternativa de evaluar los programas tiene la limitación de dejar fuera muchos aspectos importantes del funcionamiento y resultados de los programas, pero tiene la ventaja de jerarquizarlos a todos con un único índice, fácil de comprender.

Metaevaluación: selección de evaluaciones a analizar

Como fue referido al inicio de esta segunda parte, las metaevaluaciones son fundamentales para valorar la calidad de los trabajos de evaluación realizados. Dicha calidad es una de las variables que puede resultar determinante para decidir sobre la utilización de sus resultados, como se analizará posteriormente. A pesar de su importancia y de los diversos estudios propuestos por Scriven desde 1967, su desarrollo mundial ha sido lento y con énfasis en el campo educativo.³ Sólo existen avances en algunos pocos países como Reino Unido, Francia, España, Canadá, Costa Rica y México, entre otros.⁴

En el contexto mexicano, la normatividad ni siquiera las menciona. Si bien a nivel federal se ha planteado en diversas oportunidades el interés de realizar metaevaluaciones con asesoría de algún organismo internacional que analicen, valoren y contribuyan a mejorar la calidad de las evaluaciones realizadas, el Coneval no ha realizado ninguna;⁵ en cambio existen varias efectuadas en forma independiente a las evaluaciones federales.⁶ Es importante señalar que, en 2013 la Sedesol publicó trabajos que denominó como tales para cubrir las

³ A nivel teórico, el tema ha sido expuesto por Scriven (1967, 1969, 1976 y 1991), Stufflebeam (1974), House (1980; en español 1997), Schwandt y Halpern (1988), Stufflebeam y Shinkfield (1993), Díaz (2002), Díaz y Rosales (2003), Cardozo (2006, 2009, 2012a).

⁴ Una síntesis de los realizados por Lacasse (1993), Lascoumes (1998), Bustelo (2001 y 2004), Lajoudie y Toulemonde (2002), Centeno (2002), Larrú (2003), García (2003 y 2004), Cardozo (2006 y 2009) se puede consultar en Cardozo (2012a). En Reino Unido se ubicaron: a) *Local major schemes: meta evaluation*, publicada en agosto de 2014 por el Departamento del Transporte; b) *Metaevaluación de Londres 2012*, publicada en julio de 2013 por el Departamento de Cultura, Medios de Comunicación y Deportes; c) *Meta Evaluation of Mine Action and Development*, que data de julio de 2012 y fue encargada por el Departamento para el Desarrollo Internacional; y d) *Meta-evaluation of the Local Government Modernisation Agenda: Progress Report in Service Improvement in Local Government*, solicitada en 2005 por la Oficina del primer ministro y con informe final de 2007.

⁵ Entre sus mecanismos para asegurar la calidad, en la respuesta al cuestionario elaborado para la presente investigación, Coneval sólo menciona: la revisión y elaboración de TR, la definición de perfiles de evaluadores, la revisión de informes por pares y el análisis de la satisfacción de los usuarios.

⁶ Véanse Vargas (2007), Schwentesius *et al.* (2007) y Cardozo (2006, 2009 y 2012a).

evaluaciones realizadas a 20 programas en 2007-2012 (11 operados centralmente y nueve sectorizados), en el periodo 2007-2012, identificadas como evaluaciones complementarias (Cuadro 23).

CUADRO 23
“Metaevaluaciones” 2007-2012 coordinadas por la Sedesol

Programa	Institución evaluadora	Costo
Desarrollo Humano Oportunidades	Universidad A. de Chiapas	1 685 000.00
Abasto Social de Leche	Universidad A. de Chiapas	1 685 000.00
Abasto Rural	Universidad A. de Chiapas	1 685 000.00
Atención a Jornaleros Agrícolas	Universidad A. de Chiapas	1 395 000.00
70 y Más	Universidad A. de Chiapas	1 395 000.00
Coinversión Social	Universidad A. de Chiapas	1 395 000.00
Apoyo Alimentario	Universidad A. de Chiapas	1 395 000.00
Apoyo a las Instancias de Mujeres en las Entidades Federativas	Universidad A. de Chiapas	1 395 000.00
Apoyo a Vecindados en Pobreza Patrimonial para Regularizar Asentamientos Humanos Irregulares	El Colegio de la Frontera Norte	1 200 000.00
Desarrollo de Zonas Prioritarias	El Colegio de la Frontera Norte	1 200 000.00
Ahorro y Subsidio para la Vivienda Tu Casa	El Colegio de la Frontera Norte	1 200 000.00
Vivienda Rural	El Colegio de la Frontera Norte	1 200 000.00
Hábitat	El Colegio Mexiquense	276 850.00
3X1 para Migrantes	El Colegio Mexiquense	276 850.00
Rescate de Espacios Públicos	El Colegio Mexiquense	276 850.00
Prevención de Riesgos en Asentamientos Humanos	El Colegio Mexiquense	276 850.00
Opciones Productivas	CIDE	170 000.00
Empleo Temporal	CIDE	170 000.00
Estancias Infantiles para apoyar a Madres Trabajadoras	CIDE	170 000.00
Fondo Nacional para el Fomento de las Artesanías	CIDE	170 000.00

Fuente: elaboración propia con datos de la Sedesol (2012a).

El objetivo de dichas “metaevaluaciones” fue analizar los avances y logros en materia de evaluación y monitoreo, así como identificar áreas de oportunidad para los programas. Puesto que la metaevaluación tiene por objeto de análisis a las evaluaciones realizadas (no a los programas evaluados) en sus aspectos metodológicos, organizativos y éticos, la primera parte del objetivo es válida como guía para su realización, pero no la segunda. Los TR solicitaron un trabajo de gabinete utilizando los informes de evaluaciones, las RO, los ASM, las MIR, las fichas técnicas, entre otros documentos. El Coneval se reservó el derecho de emitir una posición institucional respecto de las mismas, pero se desconoce que lo haya hecho.

Cuando se revisa el contenido de sus informes, más que de una metaevaluación se trata de una síntesis analítica del contenido de evaluaciones previas que incluyen: un análisis FODA, hallazgos relevantes, propuesta de una agenda de evaluaciones futuras, mejoras a las MIR, conclusiones y recomendaciones. Algunas se refieren a los esfuerzos por utilizar las recomendaciones y darles seguimiento. Pero ninguna analiza la calidad de los trabajos de evaluación realizados: su suficiencia, rigurosidad, claridad, viabilidad de sus recomendaciones, etcétera. A lo sumo se afirma que tanto sus evaluaciones cuantitativas como cualitativas han sido rigurosas, aunque sin fundamentarlo (Oportunidades).

Existen algunas debilidades adicionales en estos trabajos organizados por la Sedesol: *a*) puesto que un metaevaluador analiza críticamente evaluaciones previas nunca puede coincidir con la persona responsable de las últimas; pero en este caso eso sucede (i.e. evaluaciones de abasto social de leche); *b*) todas las metaevaluaciones fueron concentradas en cuatro equipos de trabajo, sin mediar convocatoria abierta o por invitación (se trata de adjudicaciones directas); *c*) una metaevaluación suele requerir información adicional a la documental como la realización de entrevistas, especialmente a funcionarios y evaluadores, de cuya realización no hay indicios; y *d*) no hay ninguna justificación evidente para la diversidad de montos pagados por metaevaluación a cada equipo.

Pasando al caso del Evalúa DF, éste llevó a cabo distintos tipos de meta-evaluaciones considerando la importancia de reflexionar sobre la calidad de lo realizado: *a*) evaluación de cada informe por uno o dos consejeros, discusión y decisión colectiva sobre su aprobación, solicitud de correcciones o rechazo, así como similar proceso posterior a la aprobación para diseñar las recomendaciones que serían vinculatorias; *b*) evaluación del conjunto de

evaluaciones correspondientes a las 13 evaluaciones de programas realizadas en 2009 (Cardozo, 2012b); y c) metaevaluación de las cuatro evaluaciones de políticas de 2010 (Osuna *et al.*, 2010). La primera se realizó internamente, sin costos adicionales; mientras que la segunda fue contratada por convocatoria abierta con la Sociedad Española de Evaluación y su costo fue de \$348 000.00 (Evaluá DF, 2010a).

Finalmente, respecto del diseño de la metaevaluación de las evaluaciones a políticas y programas federales 2007-2014 que se presenta en este libro, considerando que en promedio se han evaluado más de 100 programas federales por año,⁷ se resolvió trabajar con una muestra no representativa conformada con base en los siguientes criterios: 1) incluir por lo menos una evaluación de cada uno de los tipos que se han aplicado en el periodo y dos en los más utilizados (diseño, EED y ECR); 2) considerar por lo menos dos evaluaciones de cada periodo reciente que ya haya publicado su información completa; 3) ampliar el conjunto de evaluaciones analizadas en trabajos previos, de manera de incluir no sólo las de desarrollo social sino también de otros cuatro campos relevantes: salud, educación, medioambiente y desarrollo rural; y 4) incluir por lo menos dos de cada sector. Su selección se muestra en el Cuadro 24.

El guión utilizado en el análisis de evaluaciones fue adaptado en función de los diversos tipos de evaluación aplicados a la fecha. Sin embargo, en todos los casos se analiza la suficiencia y calidad del contenido del informe en tanto producto final del proceso evaluativo, pero también los antecedentes disponibles sobre el programa y sus evaluaciones previas, el marco normativo especificado en los TR, los recursos económicos y humanos de que se dispuso, su forma de contratación y el proceso metodológico aplicado. Se excluyó para todo el conjunto el análisis del momento de la planeación y organización de las evaluaciones (tipo de evaluación a aplicar, relaciones interinstitucionales, selección del equipo de evaluadores, responsables de aprobar el informe y pagar el trabajo, etcétera) porque dicha información no se publica regularmente en la página electrónica del Coneval (salvo en algunos pocos casos sobre presupuesto y equipo evaluador), y por haberse estudiado en trabajos previos y no haber sufrido cambios significativos (Cardozo, 2006, 2009 y 2012a). En cambio, la

⁷ El inventario del Coneval identifica actualmente 278 programas sociales federales.

Selección de evaluaciones a metaevaluar a nivel federal

Periodo	Diseño	Diagnóstico	Específicas/ Específicas desempeño	Consistencia y resultados	Procesos	Indicadores	Integral Desempeño	Complementaria	Estratégica	Impacto	Total
2007-2008				Seguro Popular de Salud							1
2008-2009				Fomento Desarrollo Culturas Indígenas				Sistema Integral Calidad Salud		Oportunidades	3
2009-2010		Escuelas de Tiempo Completo				Áreas Naturales Protegidas*					2
2010-2011							Ayuda Alimentaria y Gen. Cap.			Estancias Infantiles	2
2011-2012	Manejo Áreas Naturales Protegidas			Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural							3
2012-2013	Estabilización de Cuenca y Acuíferos		Escuelas Dignas	Apoyo al Ingreso Agropecuario: Procampo					Protección Social	Escuelas de Calidad*	4
2013-2014		Escuelas de Excelencia para Abatir el Rezago									1
Total	2	1	2	3	1	1	1	1	1	3	16

* No forma parte del PAE del Coneval sino directamente realizada por la dependencia responsable del programa.

Fuente: elaboración propia, con base en la información del Coneval [http://www.coneval.gob.mx/Informes/Evaluacion], fecha de consulta: 15 de noviembre de 2014.

difusión, así como las reacciones suscitadas entre sus operadores, beneficiarios, medios de comunicación, etcétera, se analizan al final de los procesos, mientras que la utilización de resultados transformados en ASM se revisará en la tercera parte del libro. En función de toda la información previa se procede a emitir un juicio valorativo de la calidad de cada evaluación analizada.

En particular, del informe final se revisa su estructura, su claridad, los métodos aplicados, las fuentes de información utilizadas, la bibliografía consultada, la coherencia entre los resultados del análisis, las conclusiones y las recomendaciones.

En relación con el diseño de la metaevaluación a nivel del GDF, puesto que las de los dos primeros años de funcionamiento del Evalúa DF ya han sido realizadas, se cubre la totalidad de las correspondientes a 2010 y 2011, según el Cuadro 25.

Cuadro 25
Selección de evaluaciones a metaevaluar del GDF

Año	Diseño y operación de programas	Formulación e implementación de políticas	Diseño, operación y resultados de políticas	Operación, resultados e impacto de servicios	Desempeño	Total
2010		Política de equidad de género	Política de participación ciudadana			
		Política de transporte				3
2011				Servicios de Salud Materno-Infantil	Cambio Climático *	2
Total		2	1	1	1	5

* No forma parte de las evaluaciones coordinadas por Evalúa DF.

Fuente: elaboración propia con información del Evalúa DF (2010a y b y 2011) y Centro Mario Molina (2012).

CAPÍTULO XV

Metaevaluaciones en el área de desarrollo social federal

Myriam Cardozo Brum

Procesos del Programa Fomento y Desarrollo de Culturas Indígenas 2008-2009

Contexto de la evaluación

Ante la composición pluricultural del país, se crea en 2004 el Programa Fomento y Desarrollo de Culturas Indígenas (PFDCI) de la Comisión Nacional para el Desarrollo de los Pueblos Indígenas (CDI) en el marco del Programa Nacional para el Desarrollo de los Pueblos Indígenas 2001-2006, que tiene como antecedente el Programa de Fondos para la Cultura Indígena que el Instituto Nacional Indigenista (INI) operó en el periodo 1991-2003 y evaluado desde 2001.

En 2009 su objetivo era impulsar los procesos culturales de los pueblos y comunidades indígenas respetando su diversidad cultural y sus derechos, cubría 871 municipios y trabajaba en torno a tres ejes temáticos: preservación del patrimonio cultural, fomento y desarrollo de la creación artística, y apoyo a las manifestaciones culturales (Vermon, 2009).

Marco normativo

La evaluación se realiza en el marco normativo de la Ley Federal de Presupuesto y Responsabilidad Hacendaria y el Decreto del PEF para el ejercicio fiscal

2009. Según los Lineamientos, la evaluación de procesos ofrece información para mejorar la gestión, al analizar si el programa lleva a cabo su operación de manera eficaz y eficiente. Busca detectar fortalezas, debilidades, oportunidades y amenazas en el marco normativo del programa y en su estructura y funcionamiento, aportando elementos para formular estrategias que incrementen la efectividad operativa y enriquezcan el diseño de los programas.

Los TR vigentes en 2009¹ (Coneval, 2009a) señalan que la evaluación de procesos es de carácter cualitativo (aunque sin descartar trabajos cuantitativos) y busca describir y analizar la operación del programa mediante una combinación de análisis de gabinete y trabajo de campo, especificando el diagnóstico del programa y los alcances de la evaluación. Exige que el análisis de gabinete considere como mínimo los siguientes documentos:

- Normatividad aplicable (leyes, reglamentos, RO, lineamientos, manuales de procedimientos, entre otros).
- MIR del ejercicio fiscal a ser evaluado.
- Sistemas de información.
- Evaluaciones del programa.
- Documentos de trabajo institucionales e informes de avances de los aspectos susceptibles de mejora.

En cuanto al trabajo de campo señala que puede incluir observación directa, entrevistas estructuradas y semi-estructuradas, grupos focales y aplicación de cuestionarios, entre otros instrumentos; así como, considerar a los actores que intervienen en la gestión del programa (a nivel central y de las entidades federativas en que opere) y a la población beneficiaria, con base en el diseño de una muestra analítica que dé cuenta de la variabilidad de la gestión del programa. Esta última tiene que ser validada por la unidad administrativa que coordina la evaluación.

¹ Los TR de 2009 estuvieron vigentes hasta 2012 y fueron modificados en 2013. En total sólo se han realizado cinco evaluaciones de este tipo en 2009, ocho en 2010, una en 2011, nueve en 2012 y una en 2013.

Los TR de 2009 son muy amplios, pero concretamente solicitan la evaluación de la eficacia y suficiencia de los siguientes componentes y procesos, que serán a su vez los revisados en la presente metaevaluación:

- Descripción de la operación del programa
- Descripción de los procesos del programa
- Identificación y clasificación de los procesos
- Análisis y medición de atributos de los procesos
 - Planeación, programación y presupuestación
 - Difusión
 - Solicitud de apoyos
 - Selección de beneficiarios o de proyectos
 - Producción o compra de apoyos
 - Distribución de apoyos
 - Entrega de apoyos
 - Seguimiento a beneficiarios
 - Contraloría social
 - Supervisión y monitoreo
- Análisis global de la operación del programa.

Equipo evaluador y recursos económicos

En el caso del programa Fomento y Desarrollo de Culturas Indígenas la evaluación fue realizada por el Centro Regional de Investigaciones Multidisciplinarias (CRIM) de la Universidad Nacional Autónoma de México (UNAM) y contó con una coordinadora general, un evaluador principal y dos investigadores integrantes del equipo.

En este tipo de evaluación los TR sugieren un perfil para el coordinador de la evaluación y su equipo que incluye formación a nivel de licenciatura en antropología, psicología industrial, sociología administración de empresas, economía, ingeniería industrial y administración pública y posgrados en administración pública, economía, administración de empresas, investigación de operaciones y desarrollo organizacional. En el caso analizado la coordinadora

general es doctora en demografía y el evaluador principal doctor en sociología, disciplinas que aparecen sugeridas en el perfil a nivel de licenciatura, pero no entre los posgrados recomendados. Los colaboradores tienen licenciatura en comunicación y maestría en economía respectivamente; la primera de ellas tampoco está sugerida en los perfiles deseables. A pesar de tratarse de un programa sobre cultura indígena, no hay antropólogos en el equipo.

La contratación fue por adjudicación directa, sin mediar un proceso de convocatoria y selección entre candidatos. El costo del trabajo ascendió a \$1 300 000.00, mientras el programa contaba con un presupuesto de \$40 000 000.00.

Metodología de evaluación

El informe describe ampliamente el proceso de investigación, lo que es poco común en trabajos de evaluación: la estrategia general de evaluación así como la metodología utilizada, la identificación de las necesidades de información, su recolección y análisis. La selección de los espacios de trabajo de campo, que se realizó en 17 localidades de seis estados (Hidalgo, México, Michoacán, Oaxaca, Puebla y San Luis Potosí) en que se desarrollan los 18 proyectos incluidos en la muestra, se explica detalladamente. En particular se revisan todos los procesos solicitados por los términos de referencia.

Informe final

El informe de 79 páginas describe con claridad las características del programa y los resultados de las evaluaciones previas (presupuesto ejercido, logro de metas) que, aunque no incluyen una medición de impacto, permiten al equipo analizar al menos la cobertura alcanzada, así como una amplia lista de problemas identificados y las acciones realizadas para resolverlos. Posteriormente presenta el funcionamiento previsto para el programa según las RO y lo compara con el real. Los procesos identificados se comparan también entre las entidades federativas incluidas en el estudio. Se concluye con un estudio detallado de los hallazgos generales, identificando fortalezas, debilidades, problemas, soluciones y buenas prácticas por cada uno de los procesos desarrollados.

Realiza recomendaciones bien fundamentadas en cuatro aspectos: 1) una mejor alineación de los procesos del Programa con sus objetivos (mayor balance en la distribución de los proyectos y mejor documentación y resguardo de los testimonios culturales para la preservación del patrimonio); 2) la simplificación de los procesos del Programa (evitar la redundancia en el requisito de presentar la solicitud inicial sellada y firmada por las autoridades que en la práctica se cumple con la presentación de la propuesta del proyecto, participación del personal de la DFDCI en los dictámenes estatales, mayor orientación a la selección estratégica de proyectos con el máximo potencial de incidir en la creación y preservación del patrimonio cultural indígena, etcétera); 3) el aumento en la eficacia de los procesos evitando el desarrollo de propuestas que no serán apoyadas (simplificar el formato para la preparación de proyectos, establecer un sistema de seguimiento de todos los proyectos presentados, desarrollar y divulgar lineamientos generales más flexibles, etcétera) y 4) mejorar la consistencia interna y claridad de las RO y de su aplicación; todos con recomendaciones específicas para lograrlos.

Valoración de la evaluación

El informe construye descripciones completas de los procesos, tiene un buen análisis de problemas y realiza recomendaciones fundamentadas, pero hace una valoración muy positiva del programa a pesar de las muchas debilidades identificadas, para las que realiza recomendaciones tan importantes como aumentar la congruencia entre los procesos y los objetivos, simplificar los procesos e incrementar su eficacia.

También existen algunos puntos de los TR que aparecen poco cubiertos como el análisis de los recursos presupuestales, de la MIR, del sistema de información, la opinión de los beneficiarios y los mecanismos para su participación, principalmente.

Impacto de Oportunidades 2008-2009

Contexto de la evaluación

El PDHO estuvo vigente de 2001 a 2014 y constituyó el mayor esfuerzo realizado a la fecha por el gobierno federal para reducir la pobreza. Llevó a cabo diversos tipos de acciones intersectoriales de educación, salud y alimentación, centradas en las transferencias de efectivo condicionadas al cumplimiento de requisitos, con el objetivo de promover el bienestar general de las familias que viven en condiciones de pobreza alimentaria o cuyos ingresos son insuficientes para desarrollar capacidades básicas en sus integrantes (Sedesol, 2008).

Dada la relevancia de su alcance y presupuesto, así como su extensa duración, ha sido objeto de un sinnúmero de evaluaciones por parte de la Sedesol en 2001-2006 y en el marco normativo construido por Coneval de 2007 a la fecha. Dentro de este último, y al momento de aplicarse la evaluación que nos ocupa, había sido objeto en 2007 de una ECR, una de impacto de los nuevos apoyos para el consumo de energía eléctrica y gas, y una específica relativa a la calidad de algunos servicios.

La evaluación de impacto 2008-2009 correspondiente al título “Efectos de Oportunidades en áreas rurales a diez años de intervención” justifica su aplicación por el suficiente periodo transcurrido desde la creación del programa. Fue decidida por la Sedesol y el PDHO en 2007 y aprobada por el Coneval. Su informe incluye cuatro tomos, de los cuales, dada su amplitud, se seleccionaron sólo dos trabajos incluidos en el tomo 1, que es el que más se acerca a la medición de impacto.

Marco normativo

El Coneval afirma que uno de los principales retos en materia de evaluación es determinar si los programas sociales cumplen el propósito para los cuales fueron diseñados: mejorar las condiciones económicas y sociales de vida, relacionadas con el bienestar de los beneficiarios. Según el citado Consejo, las evaluaciones de impacto deben identificar con metodologías rigurosas el cambio en los indicadores de bienestar atribuibles, directa o indirectamente, a la ejecución

del programa federal, o sea, los efectos netos provocados por el programa. Su factibilidad debe ser verificada por el propio Coneval, la SHCP y la SFP, para lo cual estas instancias solicitan a todo programa que pretenda desarrollarla un análisis que la demuestre, con base en el que emiten su opinión. Dicho análisis debe incluir: antecedentes del programa, diseño de la intervención, objetivos de la evaluación, información y fuentes disponibles, y metodología aplicable.

En nuestra opinión, los aspectos mencionados deben cubrirse en cualquier tipo de evaluación y lo específico a valorar en la factibilidad de la evaluación de impacto debería ser la metodología propuesta, así como la disponibilidad de recursos para llevarla a cabo, tanto financieros como humanos, técnicos y de información.

En relación con la información se solicita la medición “antes” y “después” de las condiciones de vida de los beneficiarios del programa y de un grupo equivalente de no beneficiarios que actuará como “control” o estado contrafactual, o sea, la que requiere un estudio experimental, lo que implícitamente define la metodología que se suponía propondría el propio programa. En el apartado correspondiente a ésta se enlistan las técnicas para llevar a cabo estudios con metodologías experimentales o cuasiexperimentales;² sin prever otra posibilidad, dado que en su enfoque los dos mencionados son los únicos reconocidos como rigurosos.

En este trabajo se comparte la importancia atribuida por el Coneval a este tipo de evaluaciones, pero se considera que existen alternativas a su realización en la forma experimental en que este Consejo la concibe: se pueden realizar otras aproximaciones con base en métodos estadísticos y cualitativos que pueden aportar valiosa información en lugar de renunciar a la evaluación si el diseño experimental no es viable. No obstante, siendo el único método aceptado, en la metaevaluación será necesario verificar que efectivamente se utilicen indicadores apropiados y que se mida la diferencia “antes-después” en el grupo experimental y se deduzcan los efectos extraños causados por otras variables, con base en la evolución del grupo de control.

² Respecto de estos métodos, véase Cardozo (2012a).

Equipo evaluador y recursos económicos

El trabajo fue realizado por el CIESAS y el INSP. En el contexto del primero se designó a quien fungiría como investigadora responsable de la evaluación cualitativa, y en el segundo, a quien se hizo cargo del estudio cuantitativo.

La primera estudió la licenciatura en antropología social en la Universidad Iberoamericana, la maestría en estudios económicos y sociales y el doctorado en antropología en la Universidad de Manchester. En 2009 contaba ya con diversas experiencias previas en otras evaluaciones de Oportunidades. El segundo realizó su licenciatura en biología y su doctorado en política y administración en salud en el Massachusetts Institute of Technology (MIT). Fue director del Centro de Investigación en Evaluación y Encuestas el INSP y ha trabajado en la OMS y en el Banco Mundial.

El equipo de evaluación cualitativa estuvo conformado por seis investigadores, cuatro de ellos con doctorados en antropología, una socióloga y un economista, ambos especializados en el campo educativo. Por su parte, el equipo de evaluación cuantitativa se compuso de 12 investigadores, siete de ellos con formación en economía, varios a nivel doctoral y otros especializados en políticas públicas, educación y evaluación de programas y microeconometría, un doctor en salud y comportamiento social, otra especialista en nutrición y una doctora en ciencia social con maestrías en política comparada y en estudios del desarrollo. En el detalle de autores de capítulos de los cuatro tomos del informe aparecen autores adicionales.

Ambos equipos son de alto nivel pero se identifican problemas en su composición disciplinaria: el primer equipo adolece de especialistas en salud, nutrición y evaluación de políticas públicas, mientras el segundo presenta un exceso de economistas y un déficit de especialistas en las tres áreas sustantivas cubiertas por Oportunidades. Esto es congruente con un enfoque centrado en la medición econométrica de impactos a partir de los resultados de una encuesta, que adolece de explicaciones y de valoraciones suficientes para ser considerada como una evaluación.

La evaluación fue adjudicada en forma directa, sin mediar concurso alguno. La parte cuantitativa de la evaluación costó \$30 389 000.00 y la cualitativa, \$4 450 000.00, que se financiaron con recursos del BM. Ese monto constituyó

0.72% del presupuesto del programa en 2008, de \$48 568.35 millones a precios constantes de 2012 (Coneval, 2013b).

Metodología de evaluación

En el proceso de evaluación fueron seleccionadas metodologías diversas, lo que permite valorar la contribución de cada uno en los trabajos seleccionados: la “Evaluación de los efectos a diez años de Oportunidades en el desarrollo, educación y nutrición en niños entre 7 y 10 años de familias incorporadas desde el inicio del Programa” (capítulo I, tomo I de Behrman, Fernald, Gertler, Neufeld y Parker) es prioritariamente experimental y cuantitativa; mientras que “La vida después de Oportunidades: impacto del programa a diez años de su creación” (capítulo II, tomo I de Mercedes González de la Rocha y dos colaboradoras) es de carácter no experimental y cualitativa.

El primer estudio es provocado por la inquietud de saber si efectivamente se ha contribuido a interrumpir la transmisión intergeneracional de la pobreza. Utilizó datos de la Encuesta de Evaluación de los Hogares Rurales (Encel 1998, 2003 y 2007). La Encel 2007 se realizó con el objetivo de recolectar información con representatividad nacional, estatal y regional sobre el impacto de largo plazo de Oportunidades en zonas rurales, los efectos en la generación de capacidades básicas y su vinculación con la evaluación de los recursos, los procesos y los aspectos del contexto que afectaban los niveles de impacto alcanzados por el programa. La información se recogió a nivel de hogar, individuo, comunidad y proveedor de servicios; así como por grupos de población específicos.

Su muestra pretendió dar seguimiento a 40 000 hogares de los beneficiarios entrevistados como grupo de tratamiento en las versiones previas de la Encel desde 1998, pero resultó que, como era previsible, en su gran mayoría habían emigrado. En consecuencia, cuando fue posible, se compararon los datos de la Encel 1998 con la Encel 2007, pero en algunos casos se recurrió a identificar impactos diferenciados en función del tiempo exposición al programa, lo que resulta más limitado.

En cuanto al segundo estudio, éste realizó una revisión documental, seguida de la aplicación de técnicas etnográficas para medir el impacto en las ocupaciones,

en el tipo de inserción laboral y en los patrones de fecundidad de los ex becarios del PDHO, variables que se espera impacten la trasmisión intergeneracional de la pobreza. Si bien no se afirma que éstas sean las únicas variables determinantes, tampoco se menciona qué otras podrían ser consideradas y por qué éstas fueron las elegidas, salvo por la inclusión de una reflexión sobre la calidad de los servicios de salud y educación. Implícitamente se estaría afirmando que dichas variables explican también los cambios en los estados de salud, educación y nutrición de la población beneficiaria, factores identificados por Oportunidades como indicadores de pobreza.

El estudio cualitativo incluyó información de hogares indígenas, no indígenas y mestizos, en cuatro estados y doce microrregiones étnicamente diferentes, no necesariamente representativas de todos los cubiertos por el programa, que se habían incorporado desde el inicio al Progresá, antecesor del PDHO. En las microrregiones se incluyeron observaciones directas en los centros proveedores de servicios (escuelas, clínicas, centros y casas de salud) sitios laborales, eventos públicos (entregas de apoyos, asambleas ejidales y municipales, días de mercado) y viviendas. Se realizó observación participante y entrevistas abiertas y estructuradas a miembros de cada hogar, a los jefes de los Centros de Atención y Registro y a los promotores y vocales del programa, médicos y enfermeras, maestros y directores de escuela, autoridades y líderes locales, así como discusiones en grupos, lo que permitió la elaboración de 192 casos.

Informe final

En cuanto a la presentación, estructura y calidad, el tomo I cuenta con 284 páginas, de las cuales 43 corresponden al capítulo cuantitativo y 74 al cualitativo. Al inicio de cada estudio se señalan las limitaciones y alcances del mismo, así como la metodología utilizada. Los investigadores explicitan las debilidades del caso, especialmente el problema de no contar con los datos del grupo de tratamiento debido a la migración de los beneficiarios, por lo que tuvieron que utilizar bases de datos secundarias a la Encel 2007 como la *Encuesta nacional de ocupación y empleo* (ENOE) y la *Encuesta nacional sobre niveles de vida de los hogares* (ENNVIIH).

Los impactos estimados mediante el estudio cuantitativo se presentan por temas de educación y nutrición (no se incluye salud), sus resultados se discuten a la luz de la literatura previa y se agrega un apartado sobre las limitaciones del estudio. El análisis cualitativo permitió obtener mayores resultados confiables y demostró que existen diferencias significativas en los resultados alcanzados por Oportunidades en diversos estados y microrregiones. Sus resultados se presentaron en tres apartados: impactos en educación, en fecundidad y en ocupación. Esto implica que la cadena de causalidades planteada no estaba clara: primero fue necesario medir el impacto del programa en los niveles de educación y luego relacionar éstos con la ocupación y la fecundidad; queda por ver el impacto de estas dos en la transmisión intergeneracional de la pobreza, mediada por el mejoramiento en las condiciones de salud y nutrición que no aparecen en el estudio. Los dos estudios acaban con la correspondiente matriz FODA. La escasez de empleo se identifica como la principal amenaza para lograr el círculo virtuoso propuesto en el programa, a la que se agrega la baja calidad de los servicios analizados.

El tomo I incluye 45 recomendaciones correspondientes al conjunto de estudios incluidos (no sólo a los aquí analizados). La mayoría se concentra en educación y nutrición, si bien la calidad educativa o de los servicios de salud no depende de Oportunidades sino de otras instancias de la SEP y la SS. En concreto, el primer documento revisado recomienda, entre otros: continuar con el programa y sus evaluaciones, probar tantas mejoras potenciales como sea posible, desarrollar un enfoque programático de estimulación en los primeros años de vida, continuar probando estrategias para la suplementación de hierro para reducir la anemia, introducir incentivos con base en el desempeño para los estudiantes y maestros, fortalecer el componente nutricional del Programa, y mejorar el direccionamiento del Programa para enfocarse en las familias y niños más necesitados. Muchas estrategias son de simple continuación, otras tan generales como mejorar la focalización y algunas novedades como la estimulación temprana o los incentivos por desempeño, pero ninguna parece justificar procesos tan intensos y onerosos de evaluación de impacto.

En cuanto al segundo y mucho menos costoso, sus recomendaciones sobre el componente de Jóvenes con Oportunidades giran en torno a una mayor información y simplificación de trámites. Concretamente en materia de educación, reproducción temprana y empleo se dirigen a la SEP, la SS y la política pública

en su conjunto (en realidad, concentradas en la Sagarpa) y son bastante más amplias que las del informe cuantitativo. Entre las primeras resaltan: mantener el diferencial de becas para niñas y jóvenes, ofrecer alojamiento en albergues o transporte municipal para facilitar el acceso a las escuelas, continuar y expandir las becas del Programa Nacional de Becas para la Educación Superior (Pronabes) y Conafe, y mejorar la calidad de la enseñanza en cuanto a infraestructura, presencia de maestros, capacitación y contenidos curriculares pertinentes. En salud se sugiere: aumentar el número de médicos y evitar que haya centros que se queden sin ellos durante largas temporadas, mejorar abasto de medicamentos y disponibilidad de equipos de diagnóstico, continuar políticas de prevención, realizar campañas sobre alimentos nocivos o no nutritivos y ofrecer dietas con productos accesibles para la población, disposición de métodos anticonceptivos y talleres de educación sexual con personas capacitadas. En relación con el sector agrario y la ocupación de jóvenes se propone: mayor información sobre recursos, proyectos productivos y mecanismo de afiliación, así como evaluar su impacto. También convertir a los jóvenes ex becarios bilingües que han logrado altos niveles de escolaridad en agentes de desarrollo, disponiendo de una mayor parte de los subsidios existentes.

Como puede verse, si bien el estudio cualitativo presenta recomendaciones de simple continuación y otras generales como mejorar la calidad de la educación o aumentar el número de médicos, también ofrece propuestas más novedosas como el transporte municipal de estudiantes o el empleo de ex becarios como agentes de desarrollo.

Valoración de la evaluación

Se valora positivamente la seriedad del trabajo cuantitativo realizado; sin embargo, su calidad presenta limitaciones debidas a la elevada tasa de no respuesta en la Encel 2007 y la consecuente pérdida de representatividad nacional, por lo que los resultados sólo resultaron válidos para el conjunto de los beneficiarios encuestados.

En esta evaluación de impacto cuantitativo de Oportunidades nuestras discrepancias se plantean, no tanto con el desarrollo de los trabajos, sino con la selección del enfoque demandado por la normatividad del Coneval (análisis

atributivo) cuyas debilidades afloran siempre, y que los evaluadores no se cuestionan. Como alternativa se sugiere revisar los trabajos que proponen sustituirlo por un enfoque contributivo (Forss *et al.*, 2011). A esto se suma el alto costo implicado en este caso, si lo comparamos con otros tipos de evaluaciones posibles, incluso de impacto, como lo muestra el estudio cualitativo también revisado.

En relación con el último, se reconoce la utilidad de los casos aportados y sólo se recomienda clarificar la cadena causal que guía el análisis, verificando en campo la hipótesis de que la ocupación y los niveles de fecundidad de los ex becarios determinan su efectiva superación de la pobreza, o incluir nuevas variables que enriquezcan la explicación.

Como se recordará, Oportunidades fue uno de los programas para los que Sedesol solicitó una “metaevaluación” en el periodo 2007-2012. Si bien ésta puso énfasis en la valoración de la evolución de los programas, el informe correspondiente al PDHO incluyó una descripción de las evaluaciones que se le aplicaron y una breve valoración técnica de la calidad del trabajo. En el caso de la evaluación de impacto 2008-2009 coincide en señalar que la Encel fue adecuada como fuente de información del análisis cuantitativo:

[pero] debido a que no se contaba con información para construir grupos de control, se decidió realizar comparaciones por el tiempo de exposición al PDHO para observar el efecto sobre los beneficiarios. El problema de esta alternativa radica en que la diferencia de 20 meses de exposición para cada grupo puede ser un periodo corto [...] y por esta razón pueden no encontrarse efectos en largo plazo (Soto, 2012:26).

Agrega que las técnicas utilizadas para este modelo ya han sido probadas en trabajos similares, por lo que también las considera adecuadas. Si bien recomienda que la futura evaluación de impacto propuesta para el 2015-2016 tenga un enfoque mixto (cuantitativo-cualitativo), su valoración técnica no menciona el trabajo cualitativo realizado para la evaluación 2008-2009, revisada en este apartado.

Integral de desempeño de Ayuda Alimentaria y Generación de Capacidades 2010-2011

Contexto de la evaluación

Los Lineamientos establecen que las evaluaciones integrales de desempeño buscan contribuir a mejorar el diseño de políticas públicas y contar con una visión global 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, entre otros.

En el caso de los programas de Ayuda Alimentaria y Generación de Capacidades, la evaluación incluyó cuatro programas cuyos objetivos se vinculan al desarrollo de capacidades y la reducción de la desnutrición: el Programa de Apoyo Alimentario (PAL), el Programa de Abasto Rural (PAR), el Programa de Abasto Social de Leche (PASL) y el PDHO. El PAL y Oportunidades realizan transferencias monetarias y en especie, mientras que el PASL y el PAR efectúan sus ventas a precios subsidiados, en el primer caso dirigido a su padrón de beneficiarios y en el segundo a toda persona que compre en sus tiendas.

Marco normativo

En este caso, no hay TR generales aunque sí bastante homogeneidad en los contenidos de los informes. El conjunto de programas de Ayuda Alimentaria y Generación de Capacidades incluyó los problemas comunes al grupo, su pertinencia para resolverlos, los logros, las principales fortalezas y retos, y las recomendaciones, que resulta ser el contenido básico de este tipo de evaluaciones. En otros casos, se agrega la exposición de buenas prácticas externas en la materia.

Equipo evaluador y recursos económicos

El trabajo fue realizado por dos consultoras externas. Una de ellas es licenciada en economía por el Instituto Tecnológico Autónomo de México (ITAM) y maestra en políticas públicas por la Universidad de Harvard. Cuenta con experiencia

en diversas secretarías de la administración pública federal y desde 2008 es jefa del Departamento de Investigación de Políticas de Salud del Hospital Infantil de México Federico Gómez. Ha sido consultora en evaluación para organismos internacionales y diversos países. La otra consultora es licenciada en economía por el CIDE y cuenta con una maestría sobre políticas de financiamiento público de la Escuela de Economía y Ciencia Política de Londres. Ambas han continuado haciendo evaluaciones para Coneval, pero no se localizó información sobre alguna previa a la presente.

El equipo tiene buena formación en economía, pero no en salud y nutrición, aunque cuenta con publicaciones en estos temas. Se desconoce el costo de la evaluación y tampoco se dispone de datos referidos a la convocatoria, selección del equipo y forma de contratación.³

Metodología de evaluación

La evaluación presentada se basó en la información de las evaluaciones específicas de desempeño 2010-2011 (a pesar de que éstas no se incluyen en las referencias), las RO y otras fuentes externas de información como artículos de revistas especializadas y trabajos de las propias consultoras en el hospital infantil ya mencionado.

Informe final

A continuación se revisa la consideración de los siguientes aspectos que se esperaba fueran analizados en el informe de este tipo de evaluación:

- Diagnóstico institucional del o de los problemas propios del grupo (situación actual, evolución y prospección realizada por el gobierno federal): la evaluación afirma que los programas han sido creados y han crecido sin un

³ El Inventario de Programas y Acciones Federales de Desarrollo Social incluyó esta información entre 2009-2013, pero éste no aparece incluido.

diagnóstico institucional, lo que implica que el evaluador careció de todo documento para revisar este aspecto.

- Diagnósticos alternativos, incluyendo el análisis de las restricciones técnicas, financieras y políticas: se presenta un diagnóstico suficiente de los niveles de desnutrición y de sobrepeso y obesidad, y su relación con el nivel socioeconómico, con la población beneficiaria de Oportunidades y con la población indígena, efectuado por las consultoras a partir de la información generada por la Encuesta Nacional de Salud y Nutrición (Ensanut) 2006. Posteriormente se analizan restricciones como la evolución de los recursos y la disponibilidad de alimentos a nivel local, no así las de carácter político.
- Evolución, pertinencia y resultados esperados de la política aplicada (planes, programas, acciones y omisiones): se analizan los retos que Oportunidades intentó abordar, así como su permanencia actual, y se sostiene que el crecimiento de los programas no ha respondido a una clara estrategia ni se ha basado en una reflexión sobre la pertinencia de mantener diversos programas con objetivos similares ni sus poblaciones objetivos, coberturas, apoyos y temporalidades. En paralelo con el crecimiento de Oportunidades, se estudia la drástica reducción de los subsidios generalizados a la alimentación en el periodo 1995-2011 a través del cambio en los recursos administrados por la Compañía Nacional de Subsistencias Populares (Conasupo) y las tiendas del Sistema Social de Abasto de Distribuidoras Conasupo (Diconsa). El trabajo es completo y bien informado.
- Consistencia de los objetivos, metas, estrategias y acciones diseñadas: se comienza por revisar los objetivos de los programas, la importancia relativa de sus transferencias monetarias y su contribución a la solución del problema alimentario, para luego analizar los servicios de salud que Oportunidades y PAL exigen a sus beneficiarios y las condiciones en que éstos pueden aportar al mejoramiento de su salud. Otro punto analizado es la evolución de la población objetivo y la cobertura de los programas, así como la consistencia del conjunto por grupos étnicos y tamaño de las localidades (complementariedades, duplicidades). No se mencionan metas, pero el análisis es bastante completo.
- Descripción de las dependencias participantes, sus relaciones y sus mecanismos de coordinación interinstitucional: se menciona que no existe un diagnóstico integral de este entramado institucional, que justifique la perti-

nencia de tener varios programas ni analice sus complementariedades, por lo que no es posible analizar el punto con información del programa. El equipo evaluador podría haber tratado de suplir esta ausencia, pero no puede considerarse como una debilidad de su trabajo debido a la falta de términos de referencia que lo exijan.

- Evaluación de la eficacia, eficiencia y equidad del conjunto de acciones realizadas: se plantea que no existe una medición de los efectos de estos programas. En la medida en que la evaluación no analiza las metas tampoco puede evaluar la eficacia del programa. En cambio, contiene datos presupuestales y mediciones de cobertura que permitirían acercarse a una medición de eficiencia en cada programa. También se pudo haber dicho algo sobre la equidad, por ejemplo, mediante los cálculos efectuados de transferencias como porcentaje del ingreso laboral distribuidas por deciles (Cuadro 8 del informe) que en 2008 presentaba una distribución bastante lógica (salvo en el decil 9) pero que en 2010 se muestra mucho más distorsionada (triplicando el error de focalización en el decil 9 y generando uno nuevo e importante en el decil 10).
- Principales logros, fortalezas y retos: se refiere vagamente a logros alcanzados al pasar de subsidios generalizados a la focalización e integralidad de acciones de Oportunidades como, por ejemplo, alcanzar una redistribución de ingreso más efectiva y eficiente; pero sólo se afirma, no tiene un análisis que lo fundamente. Inmediatamente pasa a presentar una extensa lista de limitaciones y no explicita ninguna fortaleza.
- Conclusiones y recomendaciones: se concluye que los programas analizados han contribuido a la solución del problema de desnutrición en el país pero son insuficientes para resolverlo y no están atacando problemas nuevos como el sobrepeso y la obesidad, por lo que se recomienda una estrategia más amplia que incluya el mejoramiento de las viviendas, los servicios básicos, la educación, el desarrollo de mercados, las comunicaciones y transportes, entre otros, y a toda la población. Lo más recomendado es hacer otras evaluaciones y estudios: un diagnóstico y evaluación integral del conjunto de programas y también de cada uno de éstos (lo que implica suponer poca idoneidad de los existentes), una evaluación estratégica e interdisciplinaria de nutrición, una evaluación de impacto conjunto. Existen algunas recomendaciones relativas a problemas más específicos como mejorar la información y la calidad y cobertura de los servicios de salud y educación o establecer estrategias de

complementación, pero no se plantean mecanismos expresos para alcanzarlas. En general, el apartado se fundamenta en el estudio evaluativo realizado pero podría haber abarcado aspectos adicionales.

Valoración de la evaluación

El informe cuenta con 40 páginas de información y análisis presentadas con claridad. Se percibe más enfocado a hacer un buen diagnóstico del problema alimentario y de las debilidades de los programas que intentan resolverlo, que a la evaluación de sus resultados, de los que no se presentan datos que sustenten los avances afirmados. Paradójicamente por la formación de las consultoras, se tiene un mejor manejo de datos alimentarios, que de los relativos a variables económicas (costos, eficiencia) y político-administrativas (limitaciones políticas, coordinación interinstitucional).

Aunque la evaluación del conjunto no plantea una crítica expresa a las múltiples evaluaciones individuales de los cuatro programas considerados, sus extensos listados de información faltante constituyen una fuerte crítica implícita. Sin embargo, en una evaluación de este tipo sería de esperar un esfuerzo del equipo por llenar algunos vacíos como los relativos al entramado institucional o a la aplicación de criterios como eficiencia y equidad, en lugar de incluirlo todo como recomendaciones de acciones por realizar. Evidentemente, sería de mucha ayuda que los TR solicitaran y los recursos disponibles permitieran, además del análisis documental realizado, la aplicación de por lo menos entrevistas a los operadores de los programas y a sus evaluadores.

En síntesis, la evaluación se considera de buena calidad en la mayor parte del trabajo, pero resulta insuficiente en algunos aspectos importantes.

Impacto de Estancias Infantiles para apoyar a Madres Trabajadoras 2011

Contexto de la evaluación

Ante la insuficiente oferta de servicios de cuidado infantil accesibles para la población, en 2007 surge el Programa de Estancias Infantiles para Apoyar a

Madres Trabajadoras (PEI) con 11 estancias afiliadas a la Red. En 2011 operaba ya a nivel nacional. El objetivo general del programa es apoyar a las madres y padres solos para que dispongan de tiempo que les facilite la continuación de sus estudios o su integración al mercado de trabajo. Para ello las estancias necesitan garantizar una adecuada atención infantil que promueva el desarrollo de los niños y niñas.

El programa opera bajo la responsabilidad del Sistema Nacional para el Desarrollo Integral de la Familia (SNDIF) y la Sedesol; el primero supervisa y da seguimiento el área de atención y cuidado de la población infantil, y la segunda se encarga de supervisar y verificar los aspectos administrativos del PEI. Para ser beneficiarios, los padres deben contar con un ingreso menor a 1.5 salarios mínimos per cápita y los niños(as) tener entre 1 y 3 años 11 meses de edad o, en caso de discapacidad, hasta 5 años 11 meses. El PEI otorga apoyos económicos en tres modalidades: *a*) \$700.00 de apoyo en general a los padres y \$1 400.00 en caso de discapacidad; *b*) subsidio máximo de \$61 000.00 para la creación de estancias; y *c*) apoyo máximo de \$41 000.00 para incorporarse a la Red de Estancias Infantiles. En mayo de 2011, el programa beneficiaba a 249 282 padres o madres y 265 415 niños en 9 255 estancias.

El PEI contaba en 2011 con una evaluación de diseño y otra de impacto en 2007, y tres EED de 2008, 2009 y 2010, según información de la página del Coneval. Como es muy extraño que se haga una evaluación de impacto el año en que se creó el programa, se consultó el informe adjunto y se encontró que es el mismo de 2011 que se está analizando, lo que hace suponer que la de 2007 no se realizó y se incluyó por error. Con antecedentes sólo de evaluaciones de diseño y EED resulta más convincente el comentario del informe de la evaluación de impacto 2011⁴ en el sentido de que existe escasez de evidencia sobre los efectos de estos programas, especialmente en países en vías de desarrollo, lo que justifica su trabajo. También menciona que el contexto de crisis económica vivido en 2009-2011 afectó la identificación de efectos positivos del programa en el mercado laboral.⁵

⁴ Esta fue la única evaluación de impacto realizada en 2011.

⁵ Además, en junio de 2009 se produjo un incendio en la guardería ABC de Hermosillo, Sonora, que dejó 49 niños muertos y 76 heridos. A la fecha, la situación no ha sido aclarada y ningún funcionario ha sido procesado por su responsabilidad. Esto ha provocado que, a pesar

Marco normativo

El marco normativo es el mismo presentado en la evaluación de Oportunidades, también de impacto. Ante la ausencia de TR de la presente evaluación, se retoma el objetivo planteado por el propio informe que establece que el impacto será medido en “[...] variables del mercado laboral (ingreso, empleo, formalidad, estabilidad laboral) del titular beneficiario y uso del tiempo tanto del titular beneficiario como del cuidador principal del niño que reside en el hogar. Asimismo se estima el impacto en la salud, diversidad de la dieta y desarrollo infantil de los niños beneficiarios” (INSP, 2011:4).

Equipo evaluador y recursos económicos

La evaluación fue realizada en el marco del INSP por un equipo de alto nivel integrado por siete investigadores provenientes del propio instituto, la Universidad de Washington, la Universidad de California y el ITAM. Seis de ellos tienen formación a nivel de licenciatura y posgrados en economía, tanto nacionales (CIDE, Universidad de las Américas, ITAM) como extranjeros (Universidad Católica de Perú, Universidad de Carolina del Norte, Universidad Estatal de Nueva York, Escuela de Economía y Ciencia Política de Londres, Universidad de Oxford y Universidad de Stanford), y sólo una cuenta con una maestría en políticas públicas. Una séptima integrante tiene formación en psicología con doctorado en desarrollo humano.

Tres de los evaluadores cuentan con reconocimientos y trabajos a nivel internacional en organizaciones de Naciones Unidas, BID y BM. Uno de ellos tiene amplia experiencia internacional en temas de desarrollo humano, reducción de la pobreza y evaluación de programas; otro fungió como director del Centro

de los esfuerzos por avanzar en materia de transparencia y rendición de cuentas por medio de las evaluaciones, este programa se haya negado a contestar un cuestionario general diseñado para completar la información del trabajo aquí presentado, aduciendo que todo lo relativo al PEI se considera confidencial.

de Investigación en Evaluación y Encuestas el INSP. Del resto del equipo no se conoce su experiencia previa en evaluación.

La calidad del equipo es indiscutible, en cambio su composición está excesivamente inclinada hacia la economía. Sería deseable una mayor participación de psicólogos infantiles para valorar más ampliamente el desarrollo infantil logrado y de especialistas en políticas sociales capaces de analizar y explicar el impacto social del programa. También corroborar los conocimientos y experiencia en evaluación de todos los integrantes del equipo.

No hay datos disponibles sobre el proceso de su contratación; el monto pagado por la realización de este trabajo fue de \$10 000 000.00 (Coneval, 2011i).

Metodología de evaluación

En cuanto a la definición del proceso metodológico, el informe señala que por razones de rigurosidad sería de tipo experimental, para lo cual “requiere de construir un grupo de control con características observables y no observables idénticas a las del grupo que recibe el apoyo” (INSP, 2011:5), lo que, con base en la experiencia internacional acumulada, puede calificarse de utópico, a pesar de la ocurrencia a la selección aleatoria de los integrantes de cada grupo. Los propios evaluadores explican la limitación que encontraron en su aplicación y la recurrencia a un segundo método menos ambicioso (selección por pareamiento y análisis de diferencias en diferencias), que tampoco les fue posible aplicar. Finalmente, eligieron el método *pipeline*, en que la comparación se hace entre los beneficiarios y quienes están en lista de espera, lo que facilita su acceso. Los evaluadores afirman que ambos grupos difieren sólo por la diferencia de tiempo en que solicitaron su acceso (se trata de que ésta sea pequeña), lo que comprobaron contrastando sus respectivas situaciones socioeconómicas (prueba de balance). El impacto del programa se estimó mediante un modelo ordinario de mínimos cuadrados. Este diseño permite la aparición de pequeños sesgos que los evaluadores reconocen expresamente, por lo que también intentaron complementar el estudio con la comparación de las familias con diferente tiempo de exposición al programa, pero una vez más tuvieron que desecharlo por no poder cumplir con sus requisitos. Las dificultades relatadas resaltan la poca

viabilidad de estos diseños, ya mencionadas en la evaluación de impacto del PDHO, independientemente de los limitados resultados que alcanzan.

Posteriormente los evaluadores describen la muestra, aunque sin explicar el cálculo de su tamaño y selección de estados y familias, y el tipo de información que recogieron en campo por medio de una encuesta, así como la prueba de balance aplicada. Los resultados se muestran para toda la muestra y para submuestras de madres que trabajaban y que no trabajaban antes de incorporarse al Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras.

Informe final

El informe de 315 páginas comienza describiendo el crecimiento del programa, las características sociales de la población atendida y de las encargadas de las estancias, así como los resultados de la encuesta nacional de beneficiarios aplicada para monitorear el Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras.

Como resultado de la evaluación se informan impactos favorables y significativos en la probabilidad de contar con un empleo o ampliar la jornada laboral, especialmente entre las madres que no trabajaban antes, las que son minoría en el PEI, pero en el resto de las variables relativas a los padres y cuidadores los resultados no fueron significativos. En cuanto al desarrollo del niño(a), se encontraron algunos efectos positivos, aunque de magnitud moderada y sólo para algunos subgrupos; en particular, no se encontraron en su desarrollo y en la diversidad de su dieta. En cambio, se identificaron efectos negativos en la prevalencia de enfermedades. Los resultados hallados se contextualizaron en el marco de otras experiencias disponibles en la literatura (Argentina, Colombia). Nuevamente, se resaltan las limitaciones de un método que, siendo costoso, aporta tan pocos resultados.

El informe recomienda realizar un análisis de impacto por nivel socioeconómico, lo que implica reconocer que el presente trabajo no permitió saber si éste se está dirigiendo a la población más vulnerable, como se pretende. También otros estudios con modificaciones en otras variables (i.e. montos de subsidio) pero siempre con metodologías experimentales, que podrían bien sustituirse o enriquecerse con otros aportes (análisis contributivo). En términos

más concretos se incluyen recomendaciones como: dar prioridad a madres que antes no trabajaban y que actualmente son minoría, campañas de vacunación y prevención de enfermedades con la SS, promoción del desarrollo infantil como objetivo explícito, explorar alternativas para lograrlo en estancia de mejor calidad, etcétera.

El trabajo analizado se muestra técnicamente correcto; sin embargo, uno de los grandes déficits de este tipo de evaluaciones de impacto es que miden variables y sus correlaciones pero no explican los fenómenos identificados. Además, en el caso de las evaluaciones realizadas al PEI se muestra muy poca información acerca del objetivo relacionado con el desarrollo infantil. Incluso la estimulación temprana del niño(a) es valorada en el trabajo como "...una inversión especialmente rentable para aumentar la escolaridad de las personas e incluso la productividad laboral en el futuro" (INSP, 2011:17), lo que muestra la concepción economicista y reduccionista aplicada al desarrollo de niños y niñas. Otro problema es la poca contundencia de las recomendaciones como fruto de una evaluación de impacto.

Valoración de la evaluación

La evaluación es correcta en la aplicación de su metodología, sus resultados están presentados con claridad y honestidad, lo que permite valorar su contribución como limitada porque es muy poca la información nueva que aporta, manteniendo la situación de escasez de evidencia sobre los efectos de estos programas que la propia evaluación plantea al inicio para justificar su trabajo.

La crítica que esta evaluación contribuye a reafirmar tiene que ver con los Lineamientos y TR que continúan exigiendo un enfoque experimental a pesar de que éste ha dado tantas muestras de procesos caros, afectaciones éticas y resultados de poca utilidad. Esto reforzado por equipos que se han formado exclusivamente en la aplicación de esta metodología y no promueven su complementación.

Como se revisó en el capítulo anterior, este programa también fue sujeto a una "metaevaluación" en 2007-2012 a solicitud de la Sedesol. En el caso de la evaluación revisada su informe describe los hallazgos de la evaluación y sólo realiza una crítica de orden metodológico en el sentido de que no se

compararon datos objetivos “antes-después” sino que “se hizo de forma indirecta, preguntando a la madre beneficiaria que recordara su situación antes de entrar al programa” (Cejudo, 2012:31).

Estratégica de Protección Social en México 2013

Contexto de la evaluación

La seguridad social inicia en México en 1942 con la creación del IMSS, seguida por el ISSSTE en 1959, mediante un amplio sistema contributivo tripartito. Se destina a proteger a los trabajadores y sus familias de problemas de orfandad, viudez, enfermedad, desempleo y vejez. A pesar de ello, un porcentaje importante de mexicanos carecía de servicios de seguridad social, por lo que en 2004 se creó el Sistema de Protección Social (Seguro Popular) y otros programas no contributivos (Estancias Infantiles, 70 y Más, a nivel federal) para cubrir riesgos específicos. De todas formas, en 2010 el porcentaje mencionado seguía siendo de 60.7 (Coneval, 2010a). En síntesis, no ha habido un diseño preestablecido de un sistema de protección social sino que se cuenta con “[...] un conjunto fragmentado conceptual, institucional y financieramente de ‘subsistemas’ y programas dispersos [que] implica desigualdades amplias en cobertura y beneficios, genera desincentivos importantes, distorsiona los mercados laborales, y no es claro que tenga sostenibilidad financiera” (Coneval, 2013g:21-22).

El Sistema fue objeto de una ECR en 2007-2008, una de procesos administrativos y una de impacto en 2008-2009 que, si bien señalaban oportunidades de mejora, valoraban positivamente su funcionamiento global. En 2012 se resolvió aplicarle la presente evaluación estratégica, cuyos motivos y TR se desconocen.

Marco normativo

Según los Lineamientos, este tipo de evaluación se aplica a las estrategias, políticas e instituciones relacionadas con un programa o conjunto de programas. Al igual que en el caso de las evaluaciones complementarias, los TR varían en función del caso. Estos conjuntos parecen agruparse en función de diferentes

problemas interrelacionados, como podría ser salud, alimentación y educación en Oportunidades, ya que los que contribuyen a un mismo problema y buscan mejorar el diseño serán considerados en otra categoría: la evaluación integral de desempeño. Sin embargo, el ejemplo de nutrición que se presenta a continuación sólo se refiere a programas vinculados a un tipo de problema, el caso de salud se refiere a un problema y no menciona ningún programa, y el de protección social es el único que incluye problemas y programas interrelacionados. En consecuencia, no hay claridad en la definición de este tipo de evaluaciones.

La evaluación de Nutrición y Abasto se refiere a la seguridad alimentaria y los principales problemas de nutrición, así como a los programas gubernamentales que tratan de darles solución, y realizan recomendaciones.

En la relativa a mortalidad materna, los temas considerados fueron los distintos padecimientos que la pueden provocar: hipertensión, hemorragia obstétrica, aborto, sepsis, virus de inmunodeficiencia humana (VIH), problemas respiratorios, etcétera.

La evaluación de protección social incluyó: características generales de la protección social en México, riesgos asociados a la protección social (en salud, vejez, discapacidad y orfandad, asociados con el trabajo, con los ingresos insuficientes), dispersión de la oferta gubernamental y dificultades en el acceso a la protección social.

Esto supone la elaboración del guión de metaevaluación también en función de las estrategias, políticas o instituciones vinculadas al programa o conjunto de programas que se evalúen. Además, si bien es posible pensar en un contenido similar para la evaluación de estrategias y políticas, éste deber ser sustantivamente diferente si se trata de instituciones (Cunill y Ospina, 2003). Tratándose del primer caso debería cubrir tanto los contenidos mencionados para diseño, como para implementación y resultados, o sea, similar a la ECR, pero a un nivel agregado mayor que el de programa:

- Diagnóstico institucional del o de los problemas propios del grupo: situación actual, evolución y prospección realizada por el gobierno federal.
- Diagnósticos alternativos, incluyendo el análisis de las restricciones técnicas, financieras y políticas.
- Evolución, pertinencia y resultados esperados de la política aplicada.
- Planes, programas, acciones y omisiones.

- Consistencia de los objetivos, metas, estrategias y acciones diseñadas, y de su apego al marco jurídico aplicable.
- Descripción de las dependencias participantes y sus relaciones (incluir organigramas y flujogramas).
- Análisis del diseño de mecanismos de coordinación interinstitucional.
- Análisis de la consistencia entre el diseño y la implementación.
- Análisis de su implementación en función de los criterios señalados en cada caso: existencia, pertinencia, suficiencia, calidad, efectividad, entre otros.
- Sistemas de información suficientes.
- Fuentes de financiamiento y suficiencia en cantidad, calidad y oportunidad de recursos.
- Evolución de cobertura y focalización.
- Efectividad de la difusión.
- Mecanismos de participación social y comunitaria.
- Medición de eficacia, eficiencia y economía en el uso de recursos.
- Satisfacción de la población objetivo.
- Principales logros, fortalezas y retos.
- Mecanismos de rendición de cuentas.
- Conclusiones y recomendaciones.

Equipo evaluador y recursos económicos

El equipo evaluador estuvo integrado por 14 investigadores del Hospital Infantil de México y cuatro de la Conferencia Interamericana de Seguridad Social con apoyo técnico del BM y el BID. Se trata de un equipo compuesto por un subgrupo con sólida formación médica, con especialidades en pediatría, infectología, neonatología y medicina interna, con maestrías en ciencias médicas y de la salud y doctorados en salud pública y ciencias médicas. El subgrupo no médico incluye licenciados en economía, informática, actuaría y relaciones internacionales, con maestrías en políticas públicas, políticas de financiamiento público, dirección de empresas y economía, y doctorados en política y salud pública y economía. Dos de las integrantes han realizado otras evaluaciones para el Coneval, del resto se desconoce su experiencia al respecto.

No hay ninguna información disponible sobre el proceso de convocatoria, selección del equipo, forma de contratación, precio pagado, etcétera.⁶

Metodología de evaluación

El informe tampoco menciona su proceso de trabajo, sus fuentes de información ni el tratamiento metodológico de sus datos, aparentemente documentales, estadísticos e interpretativos de sus resultados.

Informe final

El informe de 161 páginas comienza cubriendo la necesidad de definir el concepto de protección social y de revisar el diagnóstico del tema; su información es tomada de diversas fuentes pero no aclara si las instituciones responsables de estos riesgos tienen un diagnóstico propio y cuál es su nivel de coincidencia. Se describen las características generales de la protección social en México en términos de su relación con los derechos sociales, su presupuesto y gasto, su cobertura, su alcance, su acceso, y se continúa con el estudio de los principales riesgos: acceso a la salud, vejez, discapacidad, orfandad y los relativos al trabajo y al ingreso. También se analiza la situación de la población en cada uno de estos aspectos y su relación con la pobreza, así como la organización institucional del sector que los atiende, sus recursos, tipos de afiliación, acceso y uso de los servicios. Este análisis permite cubrir el requisito de revisar la pertinencia, evolución, y resultados esperados de la política aplicada.

En particular es de interés el análisis realizado al conjunto de programas, acciones y omisiones federales y estatales en cada materia, el que no llega a la revisión detallada de la consistencia de sus objetivos, metas, estrategias, acciones, operación y apego al marco jurídico aplicable. Puesto que no hay una revisión explícita de la implementación no aparecen tratados procesos como la difusión

⁶ El Inventario de Programas y Acciones Federales de Desarrollo Social incluyó esta información entre 2009-2013, pero éste no aparece incluido.

de los programas, los mecanismos de participación, el nivel de satisfacción de la población o los mecanismos de rendición de cuentas; tampoco se procede a la medición de eficacia, eficiencia y economía en el uso de recursos. En cambio, ofrece una adecuada descripción gráfica de las dependencias participantes en materia de protección social y sus interrelaciones, así como de sus insuficientes mecanismos de coordinación interinstitucional; sus presupuestos, coberturas, logros y retos.

En el tratamiento específico de las pensiones para adultos mayores se ofrece información muy rica sobre el número de programas contributivos y no contributivos, de beneficiarios, montos, requisitos, etcétera, que constituyen acciones pequeñas, dispersas e inconexas que no resuelven el problema de la población a la que se dirigen. De igual manera, en torno a los riesgos asociados al trabajo se ofrece la información detallada de los programas de políticas activas de empleo. Respecto de los servicios de cuidado infantil se analiza su oferta y distribución, con diferenciales importantes de costo y calidad entre los servicios financiados de manera contributiva y no contributiva. Otro tanto sucede con los programas vinculados al derecho social al trabajo (seguro de desempleo, capacitación, etcétera), de los que se reconoce desconocer su impacto. En materia de programas relativos a incrementar los bajos niveles de ingreso (Oportunidades, Adultos Mayores, etcétera), se identifica su número, cobertura, presupuesto, gastos per cápita, etcétera, nuevamente con gran dispersión, lógicas diferenciadas (focalización/universalismo) y falta de coordinación entre los niveles de gobierno. No se perciben intentos por avanzar en la explicación de la información presentada.

En el caso de salud, el más cercano a la formación y especialización de la mayoría de los integrantes del equipo evaluador, la información es más amplia. Después de presentar información de la morbi-mortalidad de la población, se destaca la fragmentación del sistema, sus ineficiencias, problemas de acceso, y por tanto desigualdades, y se insiste especialmente en la dispersión de programas, con información detallada de su número y presupuesto por entidad federativa y su contraste con la falta de acceso a servicios. Sin embargo, en este caso surge con mayor claridad que el trabajo es exclusivamente descriptivo y de cada variable por separado, no hay un intento explicativo a partir de los datos del conjunto de variables consideradas por entidad, especialmente en los casos en

que habiendo un alto número de programas y de presupuesto se mantienen bajos los niveles de accesibilidad. Menos aún se consideran otros problemas de su funcionamiento, como los relacionados con la calidad de los servicios. O sea que, del conjunto de problemas que enfrentan los programas, se brinda muy buena información de algunos, especialmente la relacionada con el diseño del sistema y el problema de la accesibilidad, pero falta el análisis de otros y de sus interrelaciones (entre situación epidemiológica, programas, presupuestos, accesibilidad, calidad de servicios, impacto) para avanzar hacia la explicación y la evaluación del desempeño integral del sistema.

En general, el problema más identificado y bien fundamentado es la dispersión de programas, pero el tema de la protección social no se agota con ello. Al término del estudio, este problema se analiza para el conjunto de riesgos distribuidos por temas (bienestar económico, educación, salud, etcétera) y por entidad federativa, con sus respectivos presupuestos. Por último, se recomienda una mayor coordinación de esta dispersión y la medición de los efectos logrados por los mismos en la reducción de carencias y niveles de pobreza. Además se realizan 35 recomendaciones muy detalladas; a título de ejemplo se incluyen: crear una instancia que coordine la política de protección social a nivel nacional y mecanismo para su funcionamiento, establecer un sistema de gestión único con metas e indicadores claros, universalizar el derecho a la salud con énfasis en atención primaria, hacer un programa nacional de prevención de riesgos en salud financiado con impuestos generales, aprobar la portabilidad de derechos de salud geográfica e institucional, ofrecer políticas de atención a población dispersa, entre otros.

Valoración de la evaluación

Se trata de un trabajo de alta calidad por la información ofrecida en torno de la problemática y los programas vinculados a cada riesgo, que se concentra en el estudio de algunos asuntos muy importantes (dispersión de programas, desigualdad de distribución, presupuestos insuficientes, falta de coordinación intergubernamental), descuidando otros aspectos (i.e. calidad de los servicios de salud) y el tratamiento del conjunto de variables analizadas por entidad que

permita una mayor capacidad explicativa de los fenómenos estadísticamente identificados. A pesar de ellos, las conclusiones y recomendaciones se encuentran bien fundamentadas en el trabajo presentado.

Conclusiones y recomendaciones parciales

Los programas y grupos de programas cuyas evaluaciones fueron analizadas en este capítulo llevaban un número considerable de años funcionando (salvo el PEI), por lo que se justificó el tipo de evaluación aplicada: de procesos, estratégicas, integrales de desempeño e incluso de impacto. El PEI hubiera requerido un poco más de tiempo para medir y explicar adecuadamente su impacto.

Un documento básico para contribuir a evaluar la calidad de los informes lo constituyen los TR que especifican la solicitud. Sin embargo, cuatro de las cinco evaluaciones carecieron de ella (sólo disponibles en la referida a procesos), por lo que el trabajo realizado sólo valoró el contenido del informe en el marco de los Lineamientos generales y la experiencia acumulada en trabajos previos. La ausencia de TR puede constituir parte de la explicación de la limitada contribución de los informes al mejoramiento de los programas.

Los equipos evaluadores fueron de alto y muy alto nivel; en cambio se detectaron problemas en cuanto a su composición disciplinaria (énfasis en economía en detrimento de otras disciplinas necesarias para la comprensión del programa). En tres casos no se accedió a información sobre los procesos relativos a la convocatoria, selección y contratación, y en dos al monto pagado. En PDHO y Fomento y Desarrollo de Culturas Indígenas la adjudicación fue directa (sin concurso) y, si bien se justifica que el costo del primero sea superior al segundo, resulta difícil entender que lo sea tanto (26.80 veces), valorado a la luz de las aportaciones realizadas. A su vez, el PDHO costó casi 3.5 veces más que el Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras.

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 son de carácter experimental, con el objetivo de mostrar una atribución

del tipo “causa-efecto”, los que han sido ampliamente discutidos en la literatura y sólo logran parcialmente su intención.

Los informes finales están adecuadamente estructurados, su redacción es clara y su tamaño muy variable: desde 40 páginas en la evaluación integral de desempeño a 315 en una de impacto. Todos se encuentran disponibles en la página web del Coneval pero, excepto en las evaluaciones específicas recientes, éstos no se acompañan de un documento de posicionamiento institucional ante la misma, lo que sucedió en las cinco evaluaciones estudiadas. Tampoco se realiza un seguimiento de las recomendaciones por programa que se transforman en ASM ni de su cumplimiento (sólo a nivel agregado por secretarías de Estado). En el caso de la evaluación de impacto del PDHO las recomendaciones incluidas en el tomo I parecen no haber tenido incidencia en la determinación de ASM, especialmente en materia de rediseño del programa.

Se trata de trabajos de buena calidad con algunos aspectos poco cubiertos. Las limitaciones encontradas en las evaluaciones de impacto se deben fundamentalmente a la insistencia del Coneval en trabajar con modelos experimentales, costosos y poco efectivos.

Las recomendaciones que se proponen a los problemas identificados incluyen: *a)* valorar en cada caso si el programa ya ha generado efectos identificables antes de planear su evaluación de impacto (dejar al menos que transcurran tres años desde su inicio); *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; *h)* intentar evaluaciones de impacto de tipo contributivo; *i)* delimitar el número mínimo de páginas por tipo de evaluación; y *j)* informar siempre sobre la respuesta institucional a la evaluación, las recomendaciones transformadas en ASM y su cumplimiento, desglosadas para cada evaluación realizada.

CAPÍTULO XVI
Metaevaluaciones en materia educativa federal

Myriam Cardozo Brum
Talina Olvera Mejía

Específica de desempeño
de Escuelas de Tiempo Completo 2009-2010

Contexto de la evaluación

Entre los antecedentes del programa se cuenta con el Acuerdo Nacional para la Modernización de la Educación Básica (ANMEB) de 1992 y la nueva Ley General de Educación de 1993, que persiguen el fortalecimiento de los centros escolares de educación básica. La búsqueda de nuevas prácticas escolares ha llevado al diseño de diferentes programas: Escuelas de Tiempo Completo, Escuelas de Calidad, Escuela Segura, entre otras.

Las RO de Escuelas de Tiempo Completo de 2009 señalaban una justificación amplia del programa: apoyar a las familias y fortalecer su participación en la tarea educativa, es decir, atender necesidades sociales surgidas en los últimos 25 años en relación con los cambios en la estructura familiar por la incorporación de las mujeres al mercado laboral. Por su parte, el documento Bases para la Operación del Programa Nacional de Escuelas de Tiempo Completo (2008) agregaba otro elemento: la diferencia de resultados observada entre el sistema de educación pública y el privado. Finalmente, el programa busca contribuir al aprendizaje de los alumnos de escuelas públicas de educación básica mediante la extensión de la jornada escolar a seis u ocho horas diarias. Se dirige preferentemente a las escuelas que atienden a población en condiciones desfavorables en contextos

urbano-marginales, indígenas o migrantes y las que presentan bajos resultados educativos. Dio inicio en 2007 y a la fecha de la evaluación tenía cobertura en 30 entidades federativas.

Incluye seis líneas de trabajo: *a)* fortalecimiento del aprendizaje sobre los contenidos curriculares, *b)* uso didáctico de las tecnologías de la información y la comunicación, *c)* aprendizaje de una lengua adicional, *d)* arte y cultura, *e)* alimentación saludable, y *f)* recreación y desarrollo físico.

Hasta el 2010 el programa había sido objeto de una evaluación inicial de diseño y otra EED en 2008;¹ no contaba con evaluaciones de impacto.

Marco normativo

En los Lineamientos este tipo de evaluación quedó definido por exclusión: todas las no comprendidas en los tipos previos, que pueden implicar tanto trabajo de gabinete como de campo. Si bien a la fecha en que se efectuó la evaluación no existían TR, normalmente incluyen una valoración sintética del desempeño de los programas sociales que se presenta mediante un formato homogéneo, analizando el avance en el cumplimiento de sus objetivos y metas programadas, a partir de una síntesis de la información contenida en el SED y mediante el análisis de indicadores de resultados, de servicios y de gestión. Integra los siguientes temas generales que serían recogidos en el Modelo de TR elaborado para estas evaluaciones en 2012-2013 (Coneval, 2013a):

- Resultados: avance en la atención del problema o necesidad para el que fue creado el programa.
- Productos: entrega de bienes y servicios a la población beneficiaria y ejercicio del presupuesto.
- Cobertura: definición y cuantificación de la población potencial, objetivo y atendida, y localización geográfica de la primera.

¹ En la página del Coneval estas evaluaciones aparecen realizadas al Programa Nacional de Horario Extendido en Primaria.

- Seguimiento de ASM: avance de los aspectos a mejorar derivados de los resultados de evaluaciones externas previas y de las acciones definidas en el Documento de Trabajo correspondiente.

Para su revisión el guión de metaevaluación incluye los cuatro aspectos antes señalados, agregando explícitamente:

- Un análisis de eficacia en el logro de metas (R/M).
- Un estudio de la relación de eficiencia entre productos y recursos (P/R).
- La evolución de la cobertura efectiva.
- El estudio de los grados de cumplimiento de los ASM en función de que éstos afecten sólo la ejecución del programa o impliquen su reformulación, y de que esto requiera o no incrementos relevantes de presupuesto.

Equipo evaluador y recursos económicos

La evaluación fue realizada por un consultor independiente y su costo fue de \$50 496.00, mientras el programa contó con un presupuesto de aproximadamente 407 millones que corresponden a los \$457.01 millones reportados por el Coneval a precios constantes de 2012 (Coneval, 2013c). Dicho consultor tiene formación de licenciatura en ciencias políticas y administración pública por la Universidad Iberoamericana, maestría en economía por El Colegio de México y doctorado en políticas públicas en la Universidad de Chicago. Ha tenido experiencias profesionales con el BID y la SHCP. Su formación general es adecuada aunque no vinculada a temas educativos ni a la evaluación; sin embargo ha realizado una decena de estas últimas, dos de las cuales se relacionan con la Secretaría de Educación Pública.

La evaluación fue contratada después de una invitación a cuando menos tres instancias pero se desconocen los detalles del proceso de selección.

Metodología de evaluación

Tampoco existe ninguna información que explique cuál fue el proceso de investigación seguido por el evaluador. Parecería que éste se basó sólo en documentos preexistentes y que no incluyó el trabajo de campo facultativo, ya que el apartado de fuentes de información sólo menciona RO, MIR, informes trimestrales, evaluaciones de diseño y EED de 2008, avances de las acciones de mejora y otros documentos institucionales.

Informe final

El informe se publica en el formato estandarizado del Coneval y cuenta con ocho páginas. Puesto que el programa es muy reciente y la información disponible bastante limitada (de los cuatro indicadores informados, dos sólo cuentan con la meta pero no con el resultado alcanzado), el evaluador ofrece una respuesta para cada apartado, pero reitera lo mismo en varios de ellos (hallazgos relevantes, avances 2010, aspectos que el programa ha decidido realizar) o presenta gráficas que no se justifican porque sólo comparan dos valores (cobertura 2008 y 2009; escuelas incorporadas 2008 y 2009; meta y valor obtenido en cursos de formación). A continuación se revisan los cuatro aspectos cuyo tratamiento se espera cuando se realiza este tipo de evaluaciones.

- Resultados, avances, eficacia en el logro de metas: sólo se informan los resultados obtenidos conforme a dos indicadores de la MIR que cuentan con información: número de escuelas incorporadas al programa en el año y de equipos técnicos formados. El evaluador hace una crítica evidente a la MIR 2009 cuya pobreza está lejos de captar la amplitud del fin, propósito y objetivo del programa y, aunque reconoce que todos los indicadores fueron modificados en 2010, mantiene su opinión de que la MIR tiene que ser reelaborada. Por otra parte, como fortaleza del programa repite lo mencionado en el apartado de hallazgos relevantes: que la ampliación de la jornada escolar “puede incrementar las oportunidades de aprendizaje [...] en la medida que se cumplan condiciones de calidad y efectividad”. En todo caso, el condicionamiento expresado, por demás certero, vuelve muy débil

esta única “fortaleza” enunciada y hace evidente que se desconoce si hay avances sustantivos que hubieran requerido la información de, al menos, los dos indicadores que carecen de ella: las escuelas que han mejorado las oportunidades de aprendizaje de sus alumnos (éste aún no mide si efectivamente mejoraron) y la calidad de los cursos de formación impartidos.

- Servicios entregados, presupuesto ejercido, y eficiencia alcanzada: se mencionan algunas acciones realizadas dentro del programa (caracterización de las escuelas incorporadas, cursos, guías pedagógicas, informes, un concurso y documentación de mejores prácticas) pero se señala que se desconoce si éstas han tenido el efecto esperado, por lo que se recomienda valorar la realización de encuestas de satisfacción entre los beneficiarios.
- Evolución de población y cobertura: se presentan los datos de la población potencial, objetivo y atendida, en valores absolutos y porcentajes de 2009 y su comparación con 2008. Lo mismo sucede respecto de la cobertura en número de entidades, municipios, localidades y escuelas, agregando el dato de la meta planeada y de su logro con una pequeña reducción (5%) debida a recortes presupuestales. En consecuencia, el punto se considera adecuadamente cubierto.
- Seguimiento de ASM: en este apartado se plantea que el programa no está sujeto a este mecanismo, sin mayor aclaración. La revisión del correspondiente documento de Posicionamiento Institucional permitió comprender que el programa argumentó en contra de las tres recomendaciones de las evaluaciones de 2008-2009 que pretendían transformarse en ASM por considerar que uno no era claro, otro no era pertinente en ese momento y para dos de ellos se carecía de recursos presupuestales. Sin embargo, la EED señala que el programa está atendiendo las recomendaciones de las evaluaciones previas, sin ninguna fundamentación ni fuente.

Las recomendaciones de la presente evaluación incluyen: la necesidad de revisar la MIR, especialmente en su lógica vertical, de definir la población objetivo y clarificar su cuantificación, y de realizar la evaluación de impacto planeada para el 2014. Se trata de problemas de diseño y de falta de evaluaciones a profundidad, que están adecuadamente fundamentadas en el trabajo, pero nada se concluye en relación con la eficacia y eficiencia de su funcionamiento, debido a la falta de información. Además el informe señala que, si bien el Programa

expresa que no es su responsabilidad la adecuada infraestructura de las escuelas y la provisión de alimentos a los alumnos, se recomienda un estudio de diagnóstico sobre las necesidades y formas de atender estas necesidades.

Valoración de la evaluación

Evidentemente una evaluación apoyada exclusivamente en información institucional, cuando ésta se limita a unos pocos indicadores que no son capaces de medir suficientemente los avances del programa y además sólo se tiene información de la mitad de ellos, no puede ser más que un mero formalismo, plagado de repeticiones, que sólo informa del número de escuelas incorporadas, del presupuesto ejercido y de algunas acciones complementarias, pero nada dice sobre los avances del programa en ninguna de sus seis líneas de trabajo, con lo que es imposible juzgar su desempeño.

Este resultado es responsabilidad principal de la vaguedad de los Lineamientos para caracterizar este tipo de evaluación (específica), en los que se deja abierta la posibilidad de hacer un trabajo de campo que contribuiría sustantivamente a evaluar el desempeño, pero que obviamente no se puede cubrir con un presupuesto de \$50 496.00. Esta debilidad se refuerza por la falta de TR. Esto explica que el evaluador no haya podido hacer un trabajo más completo, pero no lo exime de la responsabilidad que implica aceptar hacerlo si no cuenta con la información mínima necesaria.

Impacto de Escuelas de Calidad 2012

Contexto de la evaluación

El Programa Escuelas de Calidad (PEC) inició operaciones en abril de 2001 con la idea de transformar a la política educativa hacia un modelo de gestión con enfoque estratégico de la escuela que motive el trabajo colectivo y la corresponsabilidad. En consecuencia, proporciona recursos a las escuelas para adquirir materiales y equipos, financiar procesos de capacitación y promover procesos participativos de toma de decisiones para el fortalecimiento de la gestión escolar

entre alumnos, maestros, directores y padres de familia. Concretamente, se espera que logre mejores rendimientos en el aprendizaje, una disminución de las tasas de deserción y un aumento de la asistencia a clases.

El programa es operado por la SEP a través de la Dirección General de Desarrollo de la Gestión e Innovación Educativa (DGDGIE); su Coordinación Nacional se encarga de la gestión de los aspectos financieros y operativos del programa. Tiene cobertura nacional y está dirigido a escuelas públicas de educación básica en todos sus niveles y modalidades. En 2007 atendió a 39,993 y en 2012 a 50 mil centros escolares.

El financiamiento se realiza de la siguiente manera: los gobiernos estatales aportan un monto de recursos y la SEP deposita el triple del anterior en el Fideicomiso Estatal de Escuelas de Calidad. A cada escuela se le entrega una aportación inicial de hasta \$50 000.00 en general y hasta \$70 000.00 cuando son de nuevo ingreso. Adicionalmente, las escuelas mismas obtienen recursos y el fideicomiso les entrega otro monto igual al recaudado, no pudiendo exceder de \$50 000.00 para las escuelas de nuevo ingreso y de \$30 000.00 para las que tengan más de cinco años.

En el periodo 2001-2006 se realizaron evaluaciones externas al programa por parte del CIDE, en 2007 y 2008 se realizaron dos ECR, la primera efectuada también por del CIDE y la segunda por la Facultad Latinoamericana de Ciencias Sociales (Flacso). De 2008 al 2010 hay registros de EED; en 2011 se realizó otra ECR y en 2012 se encuentra en el sitio web del PEC² la llamada “evaluación de impacto no experimental” que se analiza a continuación. Además de estas evaluaciones, se cuenta con una de impacto realizada por el Skoufias y Shapiro en 2006, por parte del Banco Mundial.

² Una de las recomendaciones de los ASM de 2009 era “realizar una evaluación de impacto en 2012”. El documento emitido en abril de 2012 sobre el Seguimiento a los ASM establecía que ese año realizarían un estudio de factibilidad y de mercado para realizar una evaluación de impacto durante el ciclo escolar 2011-2012, sujeta a disponibilidad y autorización presupuestal, por lo que la recomendación se consideraba en proceso. Sin embargo, esta evaluación de impacto no experimental no aparece en los sitios web del Coneval y la Secretaría de Educación Pública.

Marco normativo

Ya se ha mencionado el marco normativo para las evaluaciones de impacto de programas federales; sin embargo, ésta parece no ajustarse a esa normatividad, pues el informe no hace referencia al Coneval ni aparece en su página. El mismo señala que el trabajo se efectuó para la Subsecretaría de Educación Básica de la Secretaría de Educación Pública.

Dado que no existen TR ni otro documento que guíe esta evaluación, se carece de información acerca de los requerimientos establecidos para el trabajo. Por lo anterior, se retoma el objetivo citado en el informe de evaluación: “[...] identificar el impacto del PEC en indicadores de calidad de educación durante los últimos 10 años. Los indicadores analizados son tasas de deserción y reprobación, así como resultados en la prueba ENLACE”. La propuesta es adecuada a los objetivos del programa, pero faltaría agregar indicadores de asistencia a clases. Además, la medición de la calidad de la educación debiera ser más amplia que el mero seguimiento de unos pocos indicadores.

Equipo evaluador y recursos económicos

La evaluación fue realizada por C230 Consultores, que lleva a cabo trabajos de diseño, evaluación, análisis e implementación de políticas y programas públicos, así como de minería de datos³ y análisis cuantitativo.

Esta consultora cuenta con un equipo de trabajo multidisciplinario, con experiencia en diversas áreas y posgrados en administración pública y desarrollo internacional, economía, economía cuantitativa aplicada, políticas públicas y educación, entre otros. Sin embargo, no existe información concreta de los integrantes del equipo evaluador, por lo que no es posible valorar su calidad. Tampoco hay datos sobre el proceso de contratación ni sobre el monto pagado por el trabajo.

³ La minería de datos es un proceso de extracción y almacenamiento de información relevante de amplias bases de datos, el objetivo es aprovechar el valor de la información localizada y utilizar patrones preestablecidos para tener un mejor conocimiento del área en cuestión y poder tomar decisiones más confiables.

Metodología de evaluación

La evaluación consta de un estudio de gabinete y otro calificado como de campo, aunque sólo se limitó a la prueba piloto de un cuestionario. En el trabajo de gabinete se utilizaron dos métodos alternativos aplicados en estudios previos realizados al programa (Skoufias y Shapiro, 2006; y Murnane *et al.*, 2006): diferencias en diferencias y diferencias en diferencias con pareo basado en propensión a ser seleccionado. Ambos se emplean en diseños experimentales, lo que de entrada contradice el título de esta evaluación.

El propio informe señala la dificultad enfrentada para encontrar un grupo de escuelas de comparación similar al de escuelas participantes en el PEC, dado que la selección de las últimas no es aleatoria. Por lo que el grupo de tratamiento incluyó a las escuelas beneficiadas por primera vez, y el grupo de comparación a aquellas que no fueron incluidas en el programa por lo menos durante un ciclo escolar completo posterior al inicio de la ronda. Los efectos se evaluaban al término de cada ciclo escolar, identificando así los cambios en las variables de resultados académicos, pero no las causas que explican esos cambios, lo cual ha sido mencionado como una restricción en este tipo de evaluaciones.

Se informan también los niveles educativos evaluados (primaria y secundaria), pero no la selección de la muestra. En cuanto a las fuentes de obtención de datos, éstas fueron principalmente censos escolares, datos de las escuelas que han participado en el PEC y datos de ENLACE a nivel de escuela. Las mismas se consideran adecuadas para el diseño efectuado, pero insuficientes para conocer a cabalidad el impacto del programa a diez años de su implementación.

La segunda parte de la evaluación (de campo) debió avanzar en la explicación de los resultados de la anterior. Desarrolló un nuevo instrumento de medición, en forma de cuestionario, en torno a los indicadores del Modelo de Gestión Educativa Estratégica (MGEE). A diferencia del anterior, ésta detalla de manera clara y precisa el proceso metodológico seguido, mismo que constó de tres fases y tuvo un enfoque cualitativo que incluyó reuniones con la DGDGIE de la SEP para identificar los elementos que se requería evaluar, definir los objetivos de cada instrumento y obtener la autorización y las facilidades necesarias para acceder a las escuelas y realizar un estudio piloto. El informe explica el objetivo del pilotaje así como el motivo, la forma de realizarlo y el análisis multivariado que se le aplicó a sus resultados.

La primera fase incluyó la revisión de literatura y prácticas internacionales para elaborar las preguntas del instrumento y realizar su prueba piloto. En la segunda fase se seleccionó la muestra (PEC y no PEC) y se llevó a cabo el pilotaje de la herramienta, por lo que fue necesaria la vinculación con las coordinaciones estatales y escuelas. Finalmente se realizó el análisis de resultados y el ajuste y diseño del instrumento definitivo. Los instrumentos e indicadores diseñados miden la implementación del MGEE y consideran sus criterios operativos y estándares, además de las dimensiones del PEC. Las fases del estudio se muestran adecuadas y correctas para el diseño, pero éste es sólo el primer paso para iniciar la medición cualitativa de impacto del programa que, en la presente evaluación, nunca se realiza.

En cuanto a la selección de escuelas para la prueba, se describe el proceso seguido, su fundamento y duración. Se menciona que la selección fue aleatoria aunque se buscó mantener la diversidad de modalidades, tipo y turno educativo. El estudio incluyó 100 escuelas, la mayoría primarias, del Distrito Federal y los estados de México e Hidalgo, lo que es correcto para un estudio piloto que no pretende ser representativo. El informe también incluye el contenido de los cuestionarios, a quien van dirigidos (directores, maestros, padres de familia y alumnos) y el número total de personas a quienes se les aplicó (1 300).

Informe final

El informe final de evaluación consta sólo de 22 páginas y se divide en dos partes. En la primera se expone el estudio de gabinete que incluye el objetivo de la evaluación, la metodología, los resultados, las conclusiones y las recomendaciones. Los resultados de la evaluación hacen referencia a datos muy generales de evaluaciones pasadas como las ya mencionadas de Skoufias y Shapiro y Murnane *et al.*, sobre tasas de deserción, reprobación y repetición que, en algunos casos, se actualizan hasta el año 2010. La mayoría de las gráficas incluidas no se acompañan de explicaciones o comentarios. Las conclusiones son coherentes con la información presentada y, por lo mismo, muy generales como: “el PEC parece estar contribuyendo en cierta medida a explicar cambios en la tasa de aprobación y deserción de las escuelas tratadas” (C230 Consultores, 2012:13).

El informe menciona también las restricciones metodológicas enfrentadas y su consecuencia en el sentido de no poder asegurar que las diferencias observables son atribuibles exclusivamente al PEC y no a otros factores, objetivo fundamental de los diseños experimentales como el realizado. Por último, la evaluación informa que sus resultados son consistentes con los de trabajos anteriores. Aunque en apariencia el trabajo cumple con los contenidos básicos de una evaluación, su tratamiento es insuficiente y no aporta información relevante nueva.

La segunda parte del informe se refiere a la elaboración del nuevo instrumento de medición y la aplicación de su prueba piloto. Se expone amplia y detalladamente la metodología y el análisis multivariado que permitió analizar los resultados de la prueba piloto. Los resultados se presentan en una tabla que incluye las preguntas realizadas para cada dimensión y las que finalmente fueron seleccionadas para el instrumento. En esta parte se incluye también una recomendación final, congruente con el trabajo elaborado más no con la expectativa generada en el título de la evaluación, sobre la utilización de este nuevo instrumento de medición para futuras evaluaciones del PEC. Lamentablemente, no se puede realizar una valoración del instrumento producido puesto que, a pesar de que el informe menciona que se anexa al documento, éste no lo incluye.

Valoración de la evaluación

El informe no incluye datos generales como el propósito de la evaluación, la composición del equipo evaluador o el mecanismo de contratación; tampoco un apartado de contexto o antecedentes del programa, el marco teórico que sustenta el estudio y una bibliografía. Se centra en exponer resultados de un estudio de gabinete de corte experimental y no realiza la medición de impacto no experimental (cualitativo) anunciada en su título. Retoma trabajos previos y no muestra, en sentido estricto, nuevos resultados sobre el impacto del PEC. El trabajo cualitativo se limita a la elaboración de un instrumento para aplicar en futuras evaluaciones (propone iniciar en la de 2013), cuya calidad queda en duda, puesto que no se incluye dicho instrumento dentro del informe y no hay reportes de evaluaciones en 2013 o 2014 que hayan hecho uso del mismo.

La reducida extensión del informe es un reflejo de las carencias e insuficiencias de la evaluación. En cambio, se reconoce positivamente que, durante el desarrollo

de la segunda parte del trabajo, el equipo evaluador haya mantenido una estrecha colaboración con personal de la SEP y del PEC, lo que muchas evaluaciones descuidan.

Específica de Escuelas Dignas 2013

Contexto de la evaluación

El programa de Escuelas Dignas se diseña para aplicarse entre 2013 y 2018 y tiene su antecedente en el Programa Mejores Escuelas que estuvo vigente de 2007 a 2012 (Coneval, 2013d); sin embargo, en 2012 Escuelas Dignas ya contó con una ficha técnica y presupuesto (Coneval, 2012b). El programa está operado por el Instituto Nacional de la Infraestructura Física Educativa (Inifed), organismo público descentralizado de la Administración Pública Federal, y funciona en el marco de la Ley Federal de la Infraestructura Física Educativa (Infe).

Escuela Digna intenta responder al deterioro de la infraestructura provocado por el uso intensivo y edad de los inmuebles, la falta de un mantenimiento adecuado, así como por los elementos y fenómenos naturales ocurridos, entre otros. Se dirige a las comunidades de los planteles educativos públicos de tipo básico para mejorar su estado físico, con la finalidad de que éstos obtengan la certificación total o parcial, a partir del cumplimiento de siete componentes: 1) seguridad estructural y condiciones generales de funcionamiento, 2) servicios sanitarios, 3) mobiliario y equipo, 4) área de servicios administrativos, 5) accesibilidad, 6) infraestructura para la conectividad y 7) espacios de usos múltiples. Tiene cobertura en las 32 entidades federativas.

Estrictamente no cuenta con evaluaciones, menos aún externas, ya que sólo presenta fichas institucionales de monitoreo en 2012 y 2013; que incluso informan que se desconoce el presupuesto del programa previo en el periodo 2008-2011 (Coneval, 2012b y 2013d).⁴

⁴ El Programa Institucional del Inifed, publicado en el DOF el 8 de mayo de 2014, informa que en promedio el presupuesto de Mejores Escuelas fue de 2 135 millones de pesos.

Marco normativo

Como ya se mencionó a propósito de la evaluación de Escuelas de Tiempo Completo, este tipo de evaluaciones se define por exclusión, lo que genera mucha ambigüedad sobre su posible contenido, en ausencia de TR específicos. Cuatro aspectos generales que se esperaría trataran incluyen: *a*) un análisis de eficacia en el logro de metas (R/M), *b*) un estudio de la relación de eficiencia (P/R), *c*) la evolución de la cobertura efectiva, y *d*) la información y explicación del nivel de cumplimiento de los ASM. En el caso que nos ocupa dichos aspectos deben considerar los siete componentes del estado físico de los planteles educativos, ya enunciados.

Equipo evaluador y recursos económicos

Las fichas de monitoreo son elaboradas por las propias instancias que operan el programa, en este caso la Unidad de Planeación y Evaluación de Políticas Educativas de la SEP, por lo cual no existen erogaciones económicas por concepto de contratación de evaluadores externos.

Metodología de evaluación

Como ya se mencionó la ficha de monitoreo no constituye una evaluación aunque Coneval la incluya como tal, sino la presentación de un informe por parte de la SEP, con datos estadísticos y algunos comentarios breves.

Informe final

El informe se reduce a una ficha de dos páginas de extensión que da seguimiento a los resultados, población y cobertura, presupuesto, alineación con el programa sectorial y plantea un FODA simplificado y recomendaciones. Los resultados se miden con sólo dos indicadores: porcentaje de contribución al mejoramiento de la infraestructura física educativa y porcentaje de certificaciones totales o

parciales realizadas por el programa a los planteles educativos de nivel básico. No se agrega información alguna sobre las obras incluidas en el avance ni su distribución y la obtención de certificaciones implica que, al menos parcialmente, se cumplió con los componentes pero no se informa de manera desglosada. En fin, no hay manera de verificar los porcentajes informados.

En 2012 la población potencial del programa era de 68 328 inmuebles y se atendieron 3 684 (Coneval, 2012b), lo que dejaría más de 65 000 por atender, suponiendo que no surgieran nuevas necesidades. Sin embargo, la presente ficha técnica redujo a menos de la mitad la población potencial (26 930 inmuebles), sin mediar explicación alguna. A pesar de esto, la población objetivo y atendida no alcanzaba al 10% de la potencial, pero esto no se menciona, por el contrario, sólo hay referencias al cumplimiento de la meta de atención.

Su presupuesto era muy pequeño y, aunque se incrementó 43% en términos reales respecto de 2012, no alcanzaba a representar ni 1% del total del sector. Además era sólo 40% superior al promedio que Inifed informa en el caso del programa Mejores Escuelas, si bien reconoce que éste sólo permitió algunas acciones aisladas. En 2013 se habría pasado a atender menos planteles pero con más recursos, más que duplicando los montos por plantel (Inifed, 2014). Esto tampoco amerita ningún comentario en la ficha, que además ofrece una gráfica con dos escalas distintas en el mismo eje que distorsiona la representación.

En su análisis FODA identifica una fortaleza consistente en la disposición de gran cantidad de mecanismos para el seguimiento, monitoreo y evaluación de las obras; sin embargo, no menciona ninguno, lo cual es difícil de entender en un programa que durante cuatro años no tuvo evaluación alguna y desconoce incluso el presupuesto utilizado. Como debilidad se señala que carece de metodología para cuantificar su población potencial y atendida; sin embargo, los dos últimos años las ha estado informando, aunque sin mencionar cómo fueron calculadas para que los lectores puedan entender las diferencias ya señaladas entre un año y el siguiente. En cambio, Inifed (2014) presenta un diagnóstico con el número de establecimiento que requieren atención y la necesidad a cubrir. La única recomendación presentada por la evaluación es en el sentido de definir esta metodología, como si el programa no tuviera problemas presupuestales y de cobertura muy agudos. Concretamente, propone definir unidades de medida y fuentes de información sobre poblaciones, sistematizar la información, establecer los procedimientos para depurarla y actualizarla, incluir en la MIR otras acciones

que forman parte del proceso, agregar la población objetivo en el propósito y, en general, corregir y complementar las fichas técnicas.

Valoración de la evaluación

Resalta la precariedad de la información presentada, cuando hay evidencia de que el propio Inifed dispone de datos más amplios. De los temas cuyo tratamiento se esperaba, se han presentado datos sobre la eficacia en el logro de metas y la evolución de la cobertura efectiva, aunque falta su análisis y valoración crítica. No hay un estudio de eficiencia ni se da seguimiento a los ASM, pero estos últimos no existieron en 2012, ya que la evaluación correspondiente no emitió recomendaciones. De la presente se desprende un ASM en relación con la cuantificación de la población objetivo y atendida.

Diagnóstico de Escuelas de Excelencia para abatir el Rezago Educativo 2014

Contexto de la evaluación

Los problemas de rezago educativo vinculados con el abandono escolar, la baja eficiencia terminal, los resultados de las pruebas nacionales e internacionales aplicadas, han sido ampliamente documentados, pero en 2014 se solicitó por primera vez un diagnóstico vinculado a este tema debido a la creación del programa del título, coordinado por la DGDGIE de la SEP. Sin embargo, el diagnóstico se acabó en agosto, dos meses después que sus Lineamientos de Operación fueran publicados en el DOF (SEP, 2014a) y uno de que iniciara operaciones. Cuenta para 2014 con un presupuesto de \$7 567 248 270.00.

Aunque el informe no realiza ninguna mención, en realidad el programa tiene un objetivo mucho más concreto que lo que su nombre sugiere ya que sólo se dirige a las escuelas públicas de educación básica de los niveles de primaria y secundaria que se encuentran en condiciones físicas malas o muy malas con la intención de abatir el rezago educativo mediante la realización de mejoras a su infraestructura física para impactar positivamente la operación de las escuelas

y fortalecer la autonomía de su gestión, lo que evidentemente puede contribuir pero es insuficiente para lograr su objetivo y garantizar la excelencia. Además, el mejoramiento de la infraestructura es también el objetivo del programa de Escuelas Dignas.

Marco normativo

Como ya se manifestó, las evaluaciones de diagnóstico de programas nuevos están previstas en los Lineamientos y, aunque no se habían realizado hasta 2014, en 2011 el Coneval emitió un documento con los elementos mínimos del problema público que se deberían cubrir para realizar un diseño robusto del programa que los atendería y prever la asignación de presupuesto requerida (Coneval, 2011a). Se trata de analizar un problema concreto, sus causas y consecuencias, el presupuesto requerido, así como las poblaciones involucradas y realizar una propuesta para atenderlo a la luz de experiencias internacionales exitosas, con su correspondiente diseño. No se especifica si su realización será exclusivamente de gabinete o se recurrirá a levantar información en campo.

El tratamiento formal de todos estos aspectos se cubre plenamente en el diagnóstico revisado, aunque se pierde su sentido debido a que, como ya se mencionó, el diseño del programa ya se había hecho y entró en operación antes de que éste se acabara.

Equipo evaluador y recursos económicos

Ni el informe ni las páginas de la SEP y el Coneval que lo recogen ofrecen información alguna sobre la conformación del equipo evaluador, la institución responsable del trabajo, los mecanismos de asignación del contrato ni su costo. Obtener información de los últimos suele dificultarse pero los dos primeros datos se presentan normalmente en la carátula del informe, lo que en este caso no sucede.

Metodología de evaluación

Aunque no hay referencias explícitas al respecto, el trabajo parece basarse sólo en el análisis de documentos institucionales nacionales (Coneval, SEP, Inifed y planes y programas), de otros países (Uruguay, Argentina, Perú y Panamá) y de organismos internacionales (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, Unesco, y Fondo de las Naciones Unidas para la Infancia, Unicef), algunos con referencias incompletas o con errores.⁵ Aunque el tema del rezago educativo ha sido tratado en innumerables publicaciones, el informe sólo incluye un libro y dos revistas. Tampoco hay ninguna referencia que apoye metodológicamente los instrumentos utilizados en el diseño (MIR, árbol de problemas, árbol de objetivos, etcétera). Por último, no están citados entre sus fuentes los Lineamientos de Operación del programa (previamente publicados) que el informe reproduce en la parte de diseño y los menciona una sola vez, en el último apartado (impacto presupuestario).

Informe final

El informe cuenta con 38 páginas, las 15 primeras corresponden al diagnóstico del rezago educativo y las siguientes al diseño del programa. A continuación se revisan y comentan los dos puntos incluidos en la primera parte:

- *Antecedentes*: el informe resume en dos páginas los principales problemas relativos al rezago educativo, vinculándolos con la infraestructura física deficiente; sin embargo, no menciona el origen de la propuesta ni su justificación, que es lo solicitado por el documento normativo de elementos mínimos ya referido. Sorprende que dos párrafos de la página 1 del informe sean idénticos a los incluidos en la introducción de los Lineamientos de Operación del programa (publicado antes) y con la misma fuente (Unicef y Narro *et al.*).

⁵ Por ejemplo, se refiere un documento de la Contraloría General de la República sin identificar el país de que se trata, o hay error en el periodo de duración del actual PND mexicano.

- *Identificación y descripción de problema*: se ofrecen datos sobre pobreza y se la relaciona con el acceso inequitativo a la educación, abandono escolar e infraestructura física. Con datos del Censo de Escuelas, Alumnos y Maestros de Educación Básica y Especial 2013 (Cemabe), se reproduce información sobre las carencias de infraestructura (material de pisos, paredes, comedores, instalaciones deportivas, etcétera), sin informar sobre la evolución del problema, su distribución geográfica y por grupos de población, como lo solicita el documento normativo. Se presentan cuatro experiencias nacionales, advirtiendo que sus resultados no son consistentes. Por último, el árbol de problemas no presenta un claro encadenamiento de causas y efectos.

En relación con el apartado del informe referido al diseño del programa se incluyen:

- *Objetivos*: los presentados en el Árbol correspondiente no están redactados como tales⁶ sino como el negativo de los problemas identificados en el apartado anterior (situaciones sin problema), lo que no es útil si no se acompaña de la estrategia para lograr su superación; sin embargo, esta confusión está planteada desde el propio documento normativo. A continuación de la elaboración del Árbol se espera determinar y justificar los objetivos, pero el informe los presenta antes porque ya están definidos, rompiendo con la lógica del trabajo. Éstos no cubren la solicitud de vincularse con los objetivos del PND, los programas sectoriales y los convenios internacionales.
- *Cobertura*: se reproduce lo ya diseñado en los Lineamientos de Operación del programa. Para cuantificar la población objetivo se elabora un cuadro con información de Cemabe, pero ésta se refiere realmente a la población potencial (la que tiene las carencias), por lo que nunca se determina la población objetivo (la que efectivamente será atendida) ni su distribución geográfica. En lugar de expresarse sobre la frecuencia con que debe actualizarse esta información, se dice que esto dependerá de cuando se realicen nuevos censos.
- *Diseño de la intervención*: el mecanismo de operación reproduce textualmente todo lo incluido en los Lineamientos de Operación del programa (hasta los

⁶ Comenzando con un verbo en infinitivo.

ejemplos). En particular, las etapas de la intervención no guardan un orden cronológico y se organizan con base en el punto denominado Requisitos de Participación en el Programa, por lo que no incluyen las actividades internas del propio programa, y reproducen el Cuadro de Descripción y Desarrollo de los Procesos Operativos. Obviamente no inician con el diseño del programa y la publicación de sus lineamientos, que sería lo esperable, sino que reconocen implícitamente que éstos ya fueron emitidos y comienzan con la primera actividad a desarrollar por las escuelas interesadas en participar. Expresamente se informa que la MIR se replica a partir de la disponible en el portal de la SHCP y sin comentario alguno. Tampoco se mencionan los objetivos anuales de cobertura demandados en el documento normativo. Finalmente, en relación con el costo operativo del programa la evaluación se limita a informar que existe un presupuesto ya autorizado para el programa en 2014 con su correspondiente distribución por entidad federativa.

- *Presupuesto*: en lugar de referirse a las necesidades de financiamiento se presenta el monto ya aprobado y la partida en la que están etiquetados. El documento normativo solicita que se agregue la evaluación de impacto presupuestario que haya sido presentada a la SHCP; en lugar de esto el informe sólo incluye la publicación de los Lineamientos de Operación del programa.

El informe cierra sin emitir ninguna opinión ni recomendación sobre el diseño de programa ya publicado.

Valoración de la evaluación

Este es el típico trabajo realizado para cumplir con una formalidad, que no aporta ningún beneficio, por lo que cualquier precio que se haya pagado por él resulta injustificado. Su diagnóstico se limita a reproducir información de otras fuentes, por demás escasas, sin cubrir sustantivamente lo demandado en el documento normativo de elementos mínimos a ser considerados. Su árbol de problemas, una de sus pocas contribuciones, presenta problemas de lógica interna.

Salvo que se pueda pensar en aportaciones informales previas a la entrega del informe final de este trabajo (en agosto de 2014) que pudieran haber nutrido al documento de Lineamientos de Operación del programa (publicado en junio

de 2014), respetando su cronología formal es necesario concluir que el informe reprodujo textualmente parte de los Lineamientos (bastante más extensos) y sólo aportó un árbol de objetivos, también con problemas de diseño, y confundió los conceptos de población potencial y objetivo.

Si el programa ya estaba formulado, incluso su MIR, hubiera sido más útil solicitar una evaluación de diseño que hiciera al menos una valoración de su pertinencia, congruencia, alineamiento con planes y programas sectoriales, etcétera. Además, en ningún momento se menciona la posible articulación o el riesgo de duplicidad con el programa de Escuelas Dignas, cuyo objetivo es similar.

Conclusiones y recomendaciones parciales

Tres de los cuatro programas cuyas evaluaciones fueron analizadas en este grupo son de reciente creación, el más antiguo es el de Escuelas de Tiempo Completo que data de 2007 pero que llevaba sólo dos años funcionando al momento de realizarse la evaluación revisada. En consecuencia, sus evaluaciones se clasifican como de diagnóstico, específicas y específicas de desempeño, todas más sencillas que las revisadas en materia de desarrollo social. El cuarto programa es el de Escuelas de Calidad creado en 2001, al que pretendió realizarse una evaluación más compleja (de impacto), también justificada por su antigüedad.

Ninguna cuenta estrictamente con TR pero en el caso de la evaluación diagnóstica Coneval emitió en 2011 un documento en que se señalan los elementos mínimos a incluir en este tipo de trabajos. En los otros tres, al igual que en el caso de desarrollo social, sólo se valoró el contenido del informe.

Es imposible valorar la calidad de los equipos evaluadores porque: *a*) en dos casos no hay información alguna de sus composición y, en uno de ellos, ni siquiera de la institución en que se llevó a cabo el trabajo; *b*) en otro, el informe fue efectuado de manera interna por las instancias que operan el programa; y *c*) en el tercero, se trató sólo de un consultor independiente, con formación general adecuada pero no ligada a temas educativos. Salvo por la referencia de la EED de la realización de una invitación a cuando menos tres organizaciones y el monto pagado, la opacidad se repite en torno a los procesos de convocatoria, selección, contratación, y monto pagado de los otros dos casos en que se realizó una evaluación externa.

La metodología de los tres trabajos coordinados por Coneval no está explicitada en ninguno de los casos, pero se reduce al análisis documental, aun en la EED en la que los Lineamientos señalan que se puede hacer trabajo de campo. Los documentos que se refieren provienen exclusivamente de las instancias gubernamentales, salvo en la evaluación diagnóstica en que se agregan referencias de otros países y organismos internacionales. En cambio, en la evaluación de impacto se describe tanto el diseño experimental como la metodología cualitativa que se intentó aplicar.

Dos de los informes revisados se presentan en el formato preestablecido por Coneval con sólo dos y ocho páginas de extensión; el tercero corresponde a la evaluación diagnóstica y contiene 38 páginas. En este último caso la organización y redacción del trabajo es adecuada, pero lo sorprendente es que su contenido reproduce ampliamente los Lineamientos de Operación del Programa, publicados dos meses antes de la entrega del informe final. En cuanto al informe de la evaluación de impacto, consta de sólo 22 páginas de contenido, totalmente insuficientes en cantidad y calidad para la expectativa generada.

Todos los informes coordinados por el Coneval se encuentran disponibles en su página web y, en el caso de las evaluaciones específicas, éstas se acompañan de un documento de posicionamiento institucional ante la misma. El de Escuelas de Tiempo Completo sólo plantea una discrepancia con el evaluador en cuanto al mecanismo y cuantificación de la población potencial y objetivo y no responde a la crítica efectuada en la evaluación a la MIR. El de Escuelas Dignas realiza una precisión sobre los montos presupuestales manejados en el informe, de los cuales no se habría aportado la fuente utilizada, e informa que ya existe un mecanismo para cuantificar la población potencial y atendida, pero no aclara las cifras cuestionadas. Las evaluaciones no realizan el seguimiento de las recomendaciones que se transformaron en ASM ni ofrecen una explicación de ello, pero parecería que dichas recomendaciones no existieron en el caso de Escuelas Dignas y fueron rechazadas en el de Escuelas de Tiempo Completo. Por su parte, la evaluación de impacto sólo está publicada en la página web del programa correspondiente y se desconoce si tuvo algún impacto.

Se trata de trabajos de calidad limitada debido a los pocos temas cubiertos y la utilización de un reducido número de indicadores para los que se dispone de insuficiente información, por lo que aportan poco al mejor desempeño de los programas. De los tres informes de procesos coordinados por el Coneval,

el de Escuelas de Tiempo Completo es el que más se esfuerza por cumplir con las expectativas. Son evaluaciones que sirven como ejemplo de la drástica simplificación sufrida en los años recientes con la casi eliminación de las evaluaciones de impacto y el desarrollo acelerado de mínimas EED, mismas que en 2013 llegan al extremo de sustituirse por un simple informe de seguimiento de unas pocas variables, elaborado a nivel interno del organismo operador, lo que fue reconocido por el Coneval.⁷ Es importante recordar que la LGDS en su artículo 72 autoriza al Coneval a realizar evaluaciones por sí mismo o a través de uno o varios organismos independientes del ejecutor del programa, lo que no se cumple en el caso de las presentes fichas de monitoreo. Por último, el trabajo realizado para la SEP es de mala calidad porque no cumple con su objetivo ni aporta información novedosa.

Las recomendaciones propuestas para los problemas identificados son: *a*) cumplir con la obligación legal de realizar siempre evaluaciones externas; *b*) iniciar un nuevo programa recién cuando su evaluación diagnóstica haya mostrado su factibilidad y conveniencia; *c*) emitir TR para cada tipo de evaluación y adecuarlos a la problemática concreta del programa a evaluar; *d*) conformar equipos interdisciplinarios de evaluación, experimentados y *ad hoc* a cada programa; *e*) emitir siempre convocatorias, dirigidas al menos a tres instituciones; *f*) explicitar los criterios de selección de propuestas; *g*) transparentar los montos contratados; *h*) exponer siempre la metodología utilizada e incluir la realización de trabajos de campo que verifiquen y complementen la información institucional; *i*) mejorar la comunicación entre evaluadores y operadores de los programas en la mayoría de los casos; *j*) exigir un mínimo de calidad y aportaciones a los informes de evaluación, en congruencia con su tipo; y *k*) informar siempre sobre la respuesta institucional a la evaluación, las recomendaciones transformadas en ASM y su cumplimiento, desglosadas para cada evaluación realizada.

⁷ La respuesta del Coneval al cuestionario diseñado para esta investigación señala textualmente: “Sin duda, proveen de una mirada institucional sobre el desempeño del mismo”, y resalta que las fichas son elaboradas por áreas diferentes a las responsables de su operación.

CAPÍTULO XVII

Metaevaluaciones en el campo ambiental federal

Myriam Cardozo Brum
Laura Rodríguez Cardozo
Esthela Sotelo Núñez

Indicadores de mediano plazo del Programa Nacional de Áreas Naturales Protegidas 2009

Contexto de la evaluación

En México existen dos programas federales específicos de Áreas Naturales Protegidas (ANP) que constituyen los instrumentos clave de la conservación de la biodiversidad: el Programa de Trabajo (PT) vigente en 2001-2006 y el Programa Nacional de Áreas Naturales Protegidas (PNANP) en 2007-2012. En éstos se concreta la responsabilidad asumida por la Conanp, desde su creación en el año 2000, de definir una nueva política en que “[...] se posiciona al hombre en un lugar central dentro de la problemática de los ecosistemas y su biodiversidad, pues el interés gubernamental es promover y mantener las condiciones de vida digna, las oportunidades para las presentes y futuras generaciones, así como la organización y el fortalecimiento comunitario” (Conanp, 2007:5), en coherencia con los tres ejes del enfoque de la sustentabilidad (biológico, económico y social).

El PT tuvo un énfasis marcadamente social, incluyó dos pautas de tipo ambiental (18%), seis sociales (55%) y tres económicas (27%). El PNANP¹ dio continuidad al anterior y trató de atender una de las prioridades establecidas en el Plan Nacional de Desarrollo del mismo periodo en lo relativo al eje de sustentabilidad ambiental, así como al Programa Sectorial de Medio Ambiente y Recursos Naturales. No cuenta con un objetivo general explícito, sin embargo, persigue seis objetivos estratégicos que contemplan principalmente aspectos ambientales y, en menor grado, de desarrollo social y económico de las comunidades; énfasis que se confirma al analizar el contenido de sus líneas estratégicas. Esto implica que el PNANP presenta similares contenidos que el PT pero con una inversión de sus prioridades (Rodríguez y Espejel, en prensa).

A nivel internacional se reconoce cada vez más la importancia de los aspectos socioeconómicos en el éxito de las ANP, por lo tanto, éstos se incluyen en los marcos metodológicos de su evaluación. El PNANP previó, al igual que su antecesor, un proceso de evaluación de mediano término a realizarse en coordinación con el SIMEC y sus tres subsistemas (información, monitoreo y evaluación), con base en 49 indicadores.

El informe correspondiente a dicha evaluación pone de manifiesto la importancia que la Conanp le atribuye:

Está probado a nivel internacional que una buena evaluación no sólo ayuda a entender si el proyecto ha tenido impacto o no, también se puede saber la magnitud de ese impacto y quién o quiénes resultan beneficiados; así mismo puede utilizarse como una herramienta de la administración por resultados [...] para modificar el diseño y mejorar la efectividad en el tiempo [...] Una de las preguntas que motivan las evaluaciones es conocer si los programas instrumentados han logrado los resultados esperados (Conanp, 2009:1).

A continuación se metaevalúa el proceso y producto generado por la evaluación del PNANP en el marco de sus propios conceptos y objetivos, ya que no

¹ Se trata de un pequeño programa considerando su ámbito nacional y su vigencia: 33 páginas efectivas, ya que las 17 restantes sintetizan el programa previo, sus resultados y un diagnóstico sobre la propia Comisión Nacional de Áreas Naturales Protegidas.

cuenta con RO ni forma parte del PAE del Coneval, debido a que se trató de un programa especial derivado del Plan Nacional de Desarrollo.²

Marco normativo

Los tratados y acuerdos internacionales establecen algunos principios y normas (sobre soberanía, prevención, responsabilidad, etcétera), pero no existe un instrumento que defina un conjunto de derechos y obligaciones para los ciudadanos que habitan en la cercanía, explotan o visitan las Áreas Naturales Protegidas. En México, el marco legal que respalda a las ANP se encuentra establecido en la LGEEPA, que define como su objetivo general la conservación mediante el desarrollo sustentable, mientras que sus objetivos específicos se encuentran en el artículo 45, que se complementa con el artículo 47 sobre los mecanismos de participación social previstos.

La LGEEPA define al desarrollo sustentable como:

El proceso evaluable mediante criterios e indicadores del carácter ambiental, económico y social que tiende a mejorar la calidad de vida y la productividad de las personas, que se funda en medidas apropiadas de preservación del equilibrio, protección del ambiente y aprovechamiento de recursos naturales, de manera que no se comprometa la satisfacción de las necesidades de generaciones futuras (artículo 3).

Complementariamente, conceptualiza como aprovechamiento sustentable a la utilización de los recursos naturales en forma que se respete la integridad funcional y las capacidades de carga de los ecosistemas de los que forman parte dichos recursos, por periodos indefinidos (Inc. 3). La sustentabilidad incluye los aspectos ambientales, el desarrollo socioeconómico de la región y las relaciones sociales entre los seres humanos, incluso las intergeneracionales.

² Argumento expresado por el Coneval en su respuesta al cuestionario diseñado para esta investigación.

Otro ordenamiento importante para el funcionamiento de las ANP es la Ley Federal de Derechos, que establece un pago por el uso, goce o aprovechamiento turístico de las Áreas Protegidas federales en sus artículos 198, 198 A y 283 C. Debe pagarlo todo visitante del parque, excepto residentes permanentes, menores de seis años y discapacitados. Los recursos generados por este concepto se regresan íntegros a las áreas protegidas con la finalidad de apoyar la operación, el desarrollo de infraestructura, proyectos turísticos sustentables y acciones de vigilancia primordialmente.

En materia de evaluación de las ANP existe el compromiso internacional de realizarlas. A nivel nacional, se utiliza la normatividad aplicada por el Coneval a los programas incluidos en el PAE, pero ya mencionamos que éste no es el caso del PNANP. En consecuencia, la Semarnat parece haber organizado su proceso de evaluación con mucha libertad; en particular, se desconoce que haya existido un documento (del tipo de los TR) en que se especificaran los objetivos y características que la evaluación del PNANP debía tener.

Equipo evaluador y recursos económicos

El informe nunca menciona quienes integraron el equipo responsable de organizar el trabajo por parte de la Conanp (se supondría que la Dirección de Evaluación y Seguimiento), ni a quien contrató para hacer el estudio y cuánto pagaron por él.

Los participantes del primer taller organizado con objeto de la evaluación fueron 29 miembros de la Conanp y 21 invitados entre académicos, integrantes del gobierno federal, de los estatales, y de organizaciones de la sociedad civil. Los integrantes de la Conanp fueron el Comisionado Nacional y un conjunto de directores generales y directores del nivel central y de las regiones, incluida la directora de evaluación y seguimiento. No se menciona ningún otro integrante de esta Dirección, que se supondría debería estar más representada que el resto. Entre los invitados destaca el entonces Instituto Nacional de Ecología (INE), la Procuraduría Federal de Protección al Ambiente (Profepa), la Conafor, la Conabio, la Facultad de Ciencias de la UNAM, The Nature Conservancy (TNC), el Fondo

Mundial para la Conservación de la Naturaleza (WWF), Fondo Mexicano para la Conservación de la Naturaleza (FMCN) y Pronatura México AC, entre otros.

En cuanto a la posterior reunión de trabajo realizada, en ésta participó nuevamente el comisionado nacional y 24 directores generales y directores del nivel central y de las regiones y la líder del Proyecto de Cambio Climático. Nuevamente el listado sólo incluye a la responsable de la Dirección de Evaluación y Seguimiento.

Estrictamente hablando, el taller y la reunión podrían haber constituido un espacio de diálogo para discutir los resultados de una evaluación, interna o externa, y tomar decisiones al respecto, pero aquí no se menciona un equipo técnico que realice la evaluación misma, lo que resultará congruente con el posterior análisis de su informe o producto final.

Metodología de evaluación

Según el informe correspondiente, el trabajo inició con la contratación del estudio “Revisión y análisis de las metas e indicadores del PNANP 2007-2012” para analizar las fortalezas y debilidades de sus seis líneas estratégicas.

Dicho estudio sugirió un nuevo número y contenido de los indicadores, lo que fue analizado en el Taller de Revisión y Evaluación de Mediano Término sobre la Ejecución del PNANP 2007-2012, realizado en la Ciudad de México en noviembre de 2009.

Posteriormente se reporta la realización en Morelos, el 10 de diciembre de 2009, de una nueva “Reunión de trabajo para revisar, modificar, definir y/o validar las metas del grupo de indicadores del PNANP que estarán vigentes 2010-2012” para rediseñar las metas de 2010-2012, sin que en ningún momento se evaluaran las ANP, con sus avances y retos pendientes.

Informe final

La evaluación de mediano término del PNANP (2009), al igual que la del PT (2004), constituye en realidad un ejercicio de valoración y rediseño de indica-

dores y de replaneación de metas para 2010-2012, sin presentar ni analizar ninguna información de las ANP, en un documento por demás reducido.³

El referido estudio “Revisión y análisis de las metas e indicadores del PNANP 2007-2012” recomendó precisar el significado y alcance del programa, especialmente en materia de protección, manejo y restauración; cambiar la tendencia a la elaboración y aplicación de programas por la realización de acciones más concretas para obtener impactos; y verificar la congruencia interna de su diseño (misión, visión, objetivos generales, particulares, etcétera). También sugirió modificar el número y contenido de sus indicadores debido a que sólo 17 de 49 se orientaban a la medición de resultados e impactos, ya que la tendencia era a privilegiar los referidos a procesos o gestión. Esto fue analizado en el correspondiente “Taller de revisión y evaluación de medio término sobre la ejecución del Programa Nacional de Áreas Naturales Protegidas 2007-2012”, cuyo producto fue un nuevo sistema más preciso de 30 indicadores, desglosados por líneas como sigue:

Protección: 6

Manejo: 6

Restauración: 2

Conocimiento: 3

Cultura: 4

Gestión: 9

Si bien ésta no fue la distribución sugerida en el estudio, guardó el énfasis recomendado en materia de gestión, descuidando resultados e impactos. El resto del documento desglosa detalladamente los indicadores, con sus fórmulas y unidades de medida.

En síntesis, la supuesta “evaluación” no muestra ninguna información que alimente los indicadores para monitorear la situación en 2007-2009, mucho menos hubo un evaluador externo, independiente de la Conanp, que la analizara y valorara. Incluso, al término del informe, se menciona que la Dirección General

³ Se trata de tan sólo 56 páginas, de las cuales las últimas 20 incluyen un glosario, una lista de acrónimos y varias de participantes en las reuniones efectuadas.

de Evaluación y Seguimiento de la Conanp diseñará para 2010 una estrategia nacional de monitoreo, que se aplicaría en el resto del periodo.⁴

En ausencia de una verdadera evaluación de las ANP, sólo se puede hablar de evaluación de sus indicadores, proceder a analizar el contenido real del informe (un sistema de indicadores) y proponer otro para futuras evaluaciones que resulte útil en el proceso de mejorar las políticas y programas de conservación hacia el objetivo sustentable perseguido por la Comisión Nacional de Áreas Naturales Protegidas.

En relación con el conjunto de indicadores, se analizó cuáles de ellos podrán brindar información útil sobre la gestión, los resultados y el impacto en los tres ejes de la sustentabilidad, considerando que este monitoreo, si bien es necesario, no es suficiente en sustitución de una evaluación integral de recursos, diseño, ejecución, productos e impactos. En el Cuadro 26 se clasifican los indicadores retenidos en la llamada “Evaluación de mediano término” en función de los tres ejes ligados al objetivo de sustentabilidad (ambiental, social y económico) y un cuarto relacionado con el proceso de gestión para alcanzarlos.

CUADRO 26
*Análisis de líneas e indicadores propuestos
por la Evaluación de medio término del PNANP*

Línea	Indicador	A*	S*	E*	G*
Gestión	Presencia para la vigilancia, prevención y control				X
	Superficie en proceso de restauración	x			
Resultados	Porcentaje de APF con programa de monitoreo, etcétera				X
	Porcentaje de APF costeras con programa de orden ecológico				X
	Porcentaje de APF con manuales de atención a contingencia				X
	Porcentaje de APF con programa de manejo integrado del fuego en operación				X
	Porcentaje de APF con programa de monitoreo, prevención, control y saneamiento de plagas y enfermedades forestales				X
	Localidades que participan en acciones de conservación		x		

continúa...

⁴ El documento deja la impresión de que su concepto de evaluación se reduce al monitoreo de indicadores.

CUADRO 26
(continúa)
*Análisis de líneas e indicadores propuestos
por la Evaluación de medio término del PNANP*

Línea	Indicador	A*	S*	E*	G*
	Porcentaje de APF con infraestructura de apoyo a la atención del visitante				X
	Empresas turísticas comunitarias sustentables	x	x	x	
	Número de APF con cobro de derechos			x	
	Monto recaudado por año mediante el cobro de derechos			x	
	Programas de Acción para la Conservación de Especies en Riesgo (PACE) en ejecución	x			
	Porcentaje de monitoreo de especies emblemáticas en operación	x			
	Porcentaje de los proyectos de investigación registrados en las APF dedicados a conservación	x			
	Número de APF que implementan alguno de los elementos del Programa Nacional de Cultura para la Conservación		x		
	Número de actividades anuales que fomentan la conservación de las APF	x			
	Número de personas que participan en acciones de conservación		x		
	Número de acciones de conservación con participación efectiva de instancias formales e informales		x		
	Porcentaje del territorio nacional protegido				X
	Número de ANP y otra modalidades de conservación decretadas por los estados y municipios en coordinación con la Conanp				X
	Montos captados de fuentes adicionales a las fiscales				X
	Porcentaje de ANP que realizan gestiones para el pago de servicios ecosistémicos a los propietarios de los predios			x	
	Superficie de ANP que cuenta con personal que realiza acciones de conservación				X
	Superficie de tierras adquiridas para la conservación	x			X
	Número anual de nueva designaciones internacionales en ANP y otras modalidades de conservación				X
	Número de acuerdos y proyectos interinstitucionales de cooperación				X
	Porcentaje de APF con PCyM formulados				X
Impacto	Porcentaje de APF con prácticas de uso y manejo sustentable	x	x	x	
	Porcentaje de APF donde se evalúa la pérdida de ecosistemas naturales	x			

*Nota: A=ambiental, S=social, E=económico y G=gestión

Fuente: elaboración propia a partir de la Conanp (2009).

El ejercicio efectuado arrojó el siguiente resultado: los indicadores se concentran en el eje ambiental (45%), 30% en el social y 25% en el económico; miden sobre todo resultados (87%), aunque muchos también podrían formar parte del nivel de procesos de gestión, pero sólo dos indicadores (7%) miden impacto y únicamente uno lo hace en los tres ejes de la sustentabilidad.

Los funcionarios entrevistados en la Conanp manifestaron que el indicador de impacto más robusto que a la fecha se aplica en más de 30 ANP es la tasa de cambio de uso de suelo o de transformación del hábitat, puesto que ésta y las actividades mineras constituyen las principales amenazas al medio ambiente. Además, confirmaron que este monitoreo y las auditorías aplicadas por la ASF se dirigen sólo a la gestión, al cumplimiento de metas y a la realización de actividades, por lo que se confirma un abandono casi total del tema del impacto en la sustentabilidad.

En la práctica las ANP no se rigen por el PNANP, sino que cada una debería contar con un programa de manejo alineado a éste en el que se establecen las actividades, acciones y lineamientos básicos para el manejo y la administración del área. Dichos programas prevén su evaluación (Conanp, s/f:1), sin embargo tampoco es frecuente que lo hagan.

Valoración de la evaluación

El trabajo analizado no puede considerarse una evaluación efectiva de la gestión, resultados e impactos del PNANP en los términos en que los organismos internacionales esperan que se aplique para conocer lo que sucede en las ANP y su entorno inmediato, ni lo que a nivel nacional se solicita a otros programas públicos. Puesto que, al igual que la evaluación realizada al PT en 2004, ésta se enfocó sólo en revisar el sistema de indicadores, en este libro, se la caracterizó como lo que Coneval considera una evaluación de indicadores y se procedió al análisis de éstos.

De los 30 indicadores diseñados para 2009-2012, el mayor peso relativo recayó en los ambientales, lo que muestra la pérdida de interés en los aspectos sociales y económicos. Sólo dos se acercan a realizar la medición de impacto y únicamente uno recupera los tres ejes de la sustentabilidad. En consecuencia, el análisis muestra una incongruencia entre el objetivo que la LGEEPA señala para

la creación de ANP (sustentabilidad y no sólo aumento de la superficie protegida) y su equilibrada recuperación en los instrumentos de programación; así como una debilidad de los indicadores de monitoreo diseñados en materia de impacto.

Diseño del Programa de Manejo de Áreas Naturales Protegidas 2011-2012

Contexto de la evaluación

El Programa de Conservación y Manejo (PCyM) es exigido a las ANP por el artículo 65 de la LGEEPA de 1988. Su reglamento establece que el PCyM constituye “el instrumento rector de planeación y regulación que establece las actividades, acciones y lineamientos básicos para el manejo y la administración del área natural protegida respectiva” (artículo tercero, fracción 11). Para su elaboración éste les otorga un plazo de un año a partir de la fecha de la publicación de su declaratoria en el DOF. Sin embargo, en junio de 2010 se contaba sólo con 46 PCyM de las 174 ANP existentes a esa fecha (Núñez, 2012).

Ante esta situación, en 2012 fue creado el Promanp con la intención de promover la colaboración y participación para la integración o modificación de los PCyM con base en los lineamientos establecidos por la Conanp para el cumplimiento de la obligación legal mencionada. Su meta fue preparar 32 PCyM para diciembre de 2012 y su presupuesto de diez millones de pesos.

Marco normativo

Los Lineamientos señalan que los programas sociales deben someterse a evaluación dentro de su primer año de operación (capítulo 15). Coneval informa que, para facilitar y homogeneizar los resultados de este tipo de evaluaciones (de diseño), elaboró un modelo de TR consistente en 34 preguntas referentes a las características del programa; su contribución a los objetivos estratégicos; su MIR; las RO; así como las coincidencias, complementariedades o duplicidades con otros programas. En 2012 los redujo a 21 preguntas y una descripción general del programa (Coneval 2007a y s/f:2).

Equipo evaluador y recursos económicos

La evaluación fue realizada por un solo especialista (sin equipo de apoyo), con formación en economía por la UNAM, diversos diplomados y una maestría en ciencias de planeación y sistemas empresariales de la Universidad La Salle. De 1998 al 2003, se desempeñó en la Unidad Coordinadora de Asuntos Internacionales de la Semarnat, como director general adjunto para Asuntos Hemisféricos y Biodiversidad.

Coneval informa que la evaluación fue asignada por contrato, lo que normalmente implica que no hubo concurso o invitación a otros especialistas o instituciones. La evaluación costó \$102 080.00, aproximadamente 1% del presupuesto disponible del Promanp. Ante la simplicidad del programa no cabe duda de que el evaluador contó con suficiente formación, conocimiento del medio y recursos financieros para desarrollar el trabajo.

Metodología de evaluación

Puesto que se trata de un trabajo de gabinete el proceso se redujo al análisis de documentos. El informe de evaluación (Núñez, 2012) no menciona si éstos fueron suficientes y su entrega oportuna; sin embargo, puede suponerse que así fue dada la amplia lista de referencias institucionales que incluye.

Informe final

En la presente metaevaluación se considera que una adecuada revisión del diseño debería incluir las respuestas a las preguntas solicitadas en los TR, asegurando la cobertura de aspectos como:

- La relevancia y evolución de los problemas sociales atendidos por el programa, así como la eficacia de las políticas y programas aplicados en el pasado.
- La pertinencia de los objetivos y estrategias del programa para incidir en el problema social atendido.

- La identificación y cuantificación de su población potencial y objetivo.
- La consistencia del programa con los principios y objetivos de la política social, las leyes, los planes y otros programas federales, así como su contribución a los primeros y sus sinergias o duplicidades con los últimos.
- La adecuación del programa a la cultura de la población objetivo.
- La congruencia interna de sus objetivos, metas, estrategias y RO.
- La propuesta de marco lógico o MIR, incluyendo indicadores de evaluación y seguimiento.

Puesto que el Promanp incluye un solo objetivo y un solo componente (facilitar la realización de los PCyM) fue muy sencillo que la evaluación, en términos generales, cubriera todos los aspectos requeridos, aunque en algunos casos se limitara a señalar que esa información no está disponible (i.e. población potencial y objetivo). Se percibe una tendencia a afirmar el cumplimiento de cada punto aunque en su fundamentación esto quede parcialmente cuestionado (i.e. se afirma que existe información sobre los beneficiarios porque todos firman un Convenio de Concertación, pero se reconoce que no se supo si ésta se encuentra sistematizada en un padrón). Buena parte de la fundamentación realizada consiste en la reproducción de textos legales, planes y programas relativos al tema puesto que no se consultaron documentos no institucionales.

El informe identifica cinco “áreas de oportunidad” y algún error puntual en relación con la línea basal (46 PCyM ya disponibles y no cero como indica su MIR). Además señala las complementariedades y coincidencias entre programas (ya no duplicidades como solicitaban antes los TR, mucho más útiles) sin explicar cómo las define. Parecería, analizando sus conclusiones, que la coincidencia de objetivos la considera complementariedad, cuando podría ser duplicidad; en cambio, la dirección hacia las mismas poblaciones objetivos la señala como coincidencia, pero éstas sí podrían ser complementariedades. Además, en dos casos identifica la presencia simultánea de ambas relaciones (Monitoreo Biológico en ANP e Investigación Científica y Tecnología Ambiental).

Aunque algunas de estas cuestiones son importantes, la evaluación no cuestiona el contenido sustantivo del diseño del Programa, se limita a considerar que el diagnóstico lo justifica pero no se plantea otras alternativas que pudieran ser más eficaces para resolver el problema. Tampoco explica porqué el Promanp no cuenta con RO (sólo Lineamientos), tal vez porque el programa no se

dirige a alguna población para que solicite apoyos. En este sentido se percibe otro problema relativo a la definición de los beneficiarios del programa que Núñez (2012:3) identifica como personas físicas, instituciones académicas y de investigación, centros de investigación y OSC que realicen acciones o estudios de difusión, investigación, protección, conservación de los ecosistemas y su biodiversidad; sin embargo, Núñez (2012:6) menciona que el objetivo general de la evaluación es que las mencionadas organizaciones colaboren concretamente en la integración o modificación de los PCyM, por lo que serían las ANP las beneficiarias. Por otra parte, en ningún momento se considera la participación de los pobladores de las ANP o sus áreas de influencia, si bien los documentos normativos del sector dicen promover una gobernanza participativa, lo que se retoma como una de sus ventajas en la justificación teórica. Finalmente, sería bueno mencionar la particularidad que significa diseñar un programa federal cuyo único objetivo es impulsar la realización de programas locales o regionales, obligatorios por ley.

En materia de valoración del diseño del programa, el informe se limita a repetir todo lo revisado como si se tratara de un segundo resumen (el primero es el informe ejecutivo solicitado) y emite un juicio favorable pero identifica las cinco “áreas de oportunidad” para mejorar supuestamente “la estructura y gestión del programa”, ampliamente repetidas a lo largo del documento, con sus correspondientes recomendaciones. En concreto propone especificar la unidad de medida, cuantificación y metodología para definir la población potencial y objetivo, así como recopilar, sistematizar y establecer los procedimientos para depurar y actualizar la información de sus beneficiarios. También, identificar el fin del programa en el documento normativo; incluir en la MIR como actividades del programa las acciones adicionales identificadas en los Lineamientos internos; señalar el comportamiento esperado en cada uno de los indicadores de la MIR y asignarles metas sexenales, y verificar la congruencia de las metas del Componente y del segundo indicador de la Actividad respecto a lo establecido en los Lineamientos internos.

Valoración de la evaluación

Como ya se mencionó, se responden todas las preguntas de un programa muy simplificado (habría que cuestionarse si es realmente un programa) y se identifican algunos problemas menores en 33 páginas de informe, cuatro de bibliografía y 40 de anexos (varias son páginas prácticamente vacías).

El informe resulta muy repetitivo, alrededor de 80% del trabajo se encuentra incluido en el resumen ejecutivo de tres páginas. La gran pregunta que deja la metaevaluación es por qué se pagaron \$102 080.00 por evaluar el diseño de un programa tan simple, que no identificó ninguno de los pocos asuntos de fondo que el programa podría plantear.

Diseño del Programa Estabilización de Cuencas y Acuíferos 2013*Contexto de la evaluación*

Para facilitar la administración del agua subterránea, el territorio mexicano se encuentra dividido en 653 acuíferos, publicados en el DOF el 5 de diciembre de 2001. En 2010 se calculaba que alrededor de 37% (30.1 miles de millones de metros cúbicos) del volumen total concesionado para usos consuntivos,⁵ provenía de estas fuentes de agua subterránea.

Las tendencias crecientes en el consumo del agua por parte de todos los usos consuntivos, han roto paulatinamente las condiciones de equilibrio en varios de los acuíferos existentes, incurriendo en un fuerte problema de sobreexplotación. Ésta se puede definir como aquella situación en la que, por un periodo extenso, la extracción media de agua subterránea de un acuífero supera o se aproxima a la recarga media (Custodio, 2002).

⁵ El consumo humano del agua se clasifica en usos consuntivos (agrícola, abastecimiento público, industria autoabastecida y termoeléctricas) y no consuntivos (hidroeléctricas). Las tendencias mundiales de consumo entre usos consuntivos son: 86% para agricultura, 4% para uso industrial, 7.7% para generación de energía eléctrica y 2.3% para uso urbano. Conagua, “Clasificación de los usos del agua” [www.conagua.gob.mx], fecha de consulta: 30 de mayo de 2009.

En México el problema de la sobreexplotación de acuíferos tiene larga data, pero esta situación se agrava a un ritmo acelerado. En 2010, cerca de 16% (101 acuíferos) de los acuíferos se encontraba en condiciones de sobreexplotación. Los más afectados se localizan en el territorio correspondiente a la Cuenca de México, a La Laguna, y especialmente al estado de Chihuahua. Esta sobreexplotación ha ocasionado la pérdida de ecosistemas, asentamientos y agrietamientos en la superficie. A estos problemas se suman la contaminación del agua subterránea, la intrusión de salinidad en acuíferos costeros y un incremento en los costos de extracción (Conagua, 2011).

De manera general, es posible afirmar que los acuíferos asociados con los Distritos de Riego (DR) son los que presentan mayor sobreexplotación. Esta condición propicia que temas tales como el incremento en la eficiencia de riego, la rehabilitación de DR, la ampliación de la infraestructura de riego, así como la protección y el mantenimiento de la infraestructura de competencia federal, ocupen lugares prioritarios dentro de la agenda de gestión de los recursos hídricos.

Acciones como las enlistadas en el párrafo anterior constituyen uno de los cuatro temas rectores del programa más ambicioso que existe actualmente para consolidar una política sustentable de agua. Este programa, identificado como la Agenda del Agua 2030 (AA2030), señala 38 iniciativas articuladas alrededor de cuatro temas: *a*) cuencas y acuíferos en equilibrio; *b*) ríos limpios; *c*) cobertura universal de los servicios de agua y saneamientos; y *d*) asentamientos seguros (Conagua, 2011).

Es en el contexto de la AA2030 que en 2013 se crea el programa Estabilización de Cuencas y Acuíferos a cargo de Conagua, como parte de una de las líneas estratégicas para revertir el escenario tendencial deficitario del agua. En principio, este programa se presenta como una estrategia general de atención al problema del uso intensivo del agua en acuíferos sobreexplotados y cuencas con excesiva concesión de aguas superficiales. Sin embargo, sus acciones se enfocan únicamente a la realización de acciones de conservación y modernización de los Distritos de Riego.

Marco normativo

La necesidad de mantener el equilibrio y salvaguardar la capacidad de recarga de los acuíferos de la nación se expresa en el artículo 88, fracción III, capítulo 1, título tercero de la LGEEPA. Las acciones tendientes a cubrir esta necesidad son competencia del Ejecutivo Federal, según lo estipulado en la Ley de Aguas Nacionales (LAN) en su artículo 6, fracción I.

Asimismo el artículo 7 bis de la LAN declara, en su fracción XI, que la prevención de la sobreexplotación de los acuíferos es un asunto de interés público, y más adelante, en el artículo 15, fracción III, se establece el carácter obligatorio de la planificación hídrica para la gestión integrada de los recursos hídricos, la conservación de recursos naturales, ecosistemas vitales y el medio ambiente. La formulación, implantación y evaluación de la planificación y programación hídrica comprende los subprogramas específicos que permitan corregir la sobreexplotación de aguas superficiales y subterráneas.

Otro elemento que da sustento normativo a este programa es la publicación oficial de la disponibilidad media anual de las aguas nacionales, de acuerdo con la Norma Oficial Mexicana NOM-011-CNA-2000. Con base en esta última se determinó el número de acuíferos que presentan un déficit en la disponibilidad media anual de agua subterránea, justificando la creación de un programa presupuestario cuyo objetivo inicial fuera la estabilización de cuencas y acuíferos sobreexplotados.

Este programa no cuenta con RO, su funcionamiento se rige por el artículo 32 bis de la Ley Orgánica de la Administración Pública Federal (LOAPF); y por el artículo 48 de la Ley Federal de Presupuesto y Responsabilidad Hacendaria (LFPRH); 156 fracción II, 156 A y 157 de su reglamento, en materia de autorización de aquella inversión física que requiere registro en cartera, como infraestructura productiva de largo plazo (Escolero, 2013).

Además, figura en el Decreto de PEF de 2013 en el apartado Programas Federales nuevos y de modalidad K (administrativos y de apoyo). Ese mismo año, la Conagua sometió el programa a un proceso de evaluación de su diseño, en cumplimiento con los lineamientos décimo octavo, fracción 1; y vigésimo segundo, fracción 1, de los Lineamientos. En éstos se especifica que, durante su primer año de operación, se deberá llevar a cabo una evaluación de diseño de los programas nuevos. La obligatoriedad para llevar a cabo la evaluación del

diseño de este programa en 2013 queda expresada en el PAE para el Ejercicio Fiscal 2013 de los Programas de la Administración Pública Federal, en el cual se establecen los tipos de evaluación que deberían aplicarse a los programas presupuestarios ese año y se especifica además que tales evaluaciones deberán realizarse con base en el modelo de TR establecido por el Coneval (2013e).

Equipo evaluador y recursos económicos

La evaluación externa de diseño del programa estuvo a cargo de un solo consultor independiente, con doctorado en aguas subterráneas y reconocida experiencia en la investigación sobre el tema. Su perfil garantiza un amplio conocimiento técnico del mismo, aunque quizá esta evaluación se habría beneficiado con la participación de por lo menos otro especialista en temas vinculados más al análisis de políticas, que aportara una mejor articulación entre el diagnóstico técnico realizado por el consultor, con un análisis más crítico de la MIR, basado en las entrevistas realizadas a las áreas responsables del programa. El procedimiento de contratación fue por adjudicación directa y el costo total de la evaluación ascendió a \$92 220.00. El estudio se financió con recursos fiscales de la Conagua, correspondientes al ejercicio 2013 (Escolero, 2013; Anexo 11).

Metodología de la evaluación

Con respecto al abordaje metodológico, el informe final (Escolero, 2013) sólo menciona que la evaluación se efectuó con base en entrevistas a funcionarios de la Conagua y a la documentación que éstos entregaron, en particular la MIR. En el documento se menciona que la principal fuente de información para esta evaluación son las respuestas ofrecidas por los informantes a 19 preguntas realizadas, organizadas en cinco temas: *a*) justificación de la creación y del diseño del programa, *b*) contribución del programa presupuestario a los Cinco Ejes para lograr una Democracia de Resultados, *c*) población potencial y objetivo, *d*) MIR, y *e*) complementariedades y coincidencias con otros programas federales. Las preguntas realizadas a los funcionarios de la Conagua corresponden a los TR elaborados por el Coneval en 2011.

Al ser una evaluación de diseño, se considera que la aproximación metodológica –basada en entrevistas semiestructuradas y trabajo de gabinete– fue adecuada; sin embargo, no se menciona el número de entrevistas realizadas, ni la lista de informantes clave y los criterios de selección de éstos, lo cual contribuiría a hacer más transparente el informe en términos de cuáles son las áreas técnicas cuya opinión está vertida en la evaluación, permitiendo la identificación de posibles espacios de oportunidad para la incorporación de enfoques complementarios o quizá más críticos.

Informe final

El producto de esta evaluación es un informe final de 39 páginas, estructurado en siete apartados, un resumen ejecutivo, una introducción, una sección de conclusiones y 14 anexos. Los apartados 1, 2 y 5 presentan el análisis del programa en un contexto amplio.⁶ Los apartados 3 y 4 analizan elementos específicos del diseño (población potencial y objetivo, y Matriz de Indicadores para Resultados). Por último, se incluyen dos apartados más en los que se hace una valoración del programa evaluado en términos de su diseño y de los resultados de un análisis FODA. Se considera que esta valoración es congruente con el desarrollo del ejercicio evaluativo. Es decir, la información que se incluye en la matriz FODA corresponde y se sustenta con el contenido del informe final, que se analiza a continuación:

- Identificación del problema y justificación del programa

El informe sostiene que el programa evaluado es relevante y su existencia, justificada. Para ello, analiza la información ofrecida por la entidad responsable (Conagua) con respecto a la problemática en la cual se pretende incidir.

⁶ El apartado 1 analiza la justificación de su creación; el 2 enuncia de forma esquemática la relación del programa evaluado con las metas sexenales de desarrollo; y el 5 identifica puntos de confluencia con otros programas federales.

Al respecto, el documento encuentra que el problema público que trata de atenderse (i.e. los acuíferos sobreexplotados y cuencas “sobreconcesionadas”) está claramente identificado por la Conagua y adecuadamente sustentado en el programa. Para sostener esta afirmación, se hace referencia a las publicaciones oficiales disponibles, que reportan anualmente el número de acuíferos sobreexplotados reconocidos. Se asume también que la Conagua tiene bien identificado territorialmente el problema (no sólo se conoce cuales son, sino donde están los acuíferos sobreexplotados) (Conagua, 2012).

En cuanto a las causas del problema, el informe señala la articulación de dos circunstancias a partir de las cuales un acuífero incurre en sobreexplotación (i.e. si se extrae un volumen de agua mayor que el concesionado, y/o existe una extracción de agua mayor a la recarga). Sin embargo, no se aclara si en el programa se ofrece un diagnóstico que las identifique y reconozca como justificación para su creación, o si es más bien una explicación que ofrece el propio informe para decir que el programa es necesario. Dentro del mismo listado de causas, el informe de evaluación incluye algunos aspectos que, más que explicar la existencia del problema, son la justificación del programa ofrecida por el evaluador (v.g. que una parte importante de la recarga de los acuíferos proviene de la infiltración del riego, o que el abatimiento de los niveles del agua en acuíferos sobreexplotados está ocasionando un incremento en los costos de producción en el sector agrícola, entre otros).

El informe concluye que la creación del programa está plenamente justificada, argumentando que “se requiere urgentemente llevar a cabo acciones que contribuyan a mejorar la productividad en el sector agrícola a partir de un uso eficiente y sustentable del agua” (Escolero, 2013:28). Se considera que la evaluación logra dar cuenta no sólo de la necesidad y relevancia del programa, sino de aquellos elementos que éste aporta como evidencia del problema a atender. No obstante, el análisis sobre la justificación del programa incluido en el informe no menciona si los objetivos y estrategias del programa evaluado son, en efecto, los mecanismos más *ad hoc* para incidir sobre las causas identificadas.

- Contribución del programa presupuestario al PND

El informe incluye un apartado sobre cómo este programa se alinea a las cinco “Metas nacionales para lograr una democracia de resultados”, establecidas en el PND 2013-2018. La información que ofrece es escueta, enunciando solamente la meta, objetivo, estrategia y línea de acción del PND a los que el programa se suscribe. Se considera que en este apartado debió incluirse, además de la necesaria alineación con el sistema de metas sexenales, la consistencia del programa evaluado con los principios y objetivos de la política social en general y las posibles contribuciones que éste podría tener en el cumplimiento de tales objetivos.

- Identificación y cuantificación de su población potencial y objetivo

En el informe se señala que el programa tiene definidas sus poblaciones (potencial y objetivo) y que las definiciones cuentan con fuentes de información disponibles. Sin embargo, inmediatamente después se afirma que el mismo no cuenta con un documento oficial donde se especifique cual es la población potencial y la población objetivo. Tampoco se cita ningún documento de diagnóstico del problema que especifique unidades de medida y metodologías para su definición.

Las poblaciones que se señalan en el informe son más bien definidas por el evaluador, considerando: *a*) que el programa se enfoca en acciones de conservación y modernización en los DR; y *b*) que el problema que busca atender el programa es el de la sobreexplotación de acuíferos. Con estos dos elementos, en el informe “se propone como criterio para definir a la población objetivo a aquellos Distritos de Riego localizados dentro de los límites geográficos de acuíferos sobreexplotados o cuencas con sobreconcesión de aguas superficiales” (Escolero, 2013:13).

A este respecto, si bien se espera que una evaluación de diseño dé como resultado la identificación de ventanas de oportunidad para mejorar el programa evaluado, consideramos que las propuestas de mejora no deberían usarse como información ya disponible en el programa, ni como sustento para responder la guía de preguntas sugerida en los TR del Coneval. Incluso, en el propio informe final se reserva uno de los anexos —el Anexo 2— para exponer la metodología sugerida para identificar y cuantificar a la población potencial

y a la población objetivo, por lo que la respuesta a la pregunta expresa de si las poblaciones (potencial y objetivo) están definidas en documentos oficiales y/o en el diagnóstico del problema, debería haber sido negativa.

En el mismo sentido, el informe responde afirmativamente a la pregunta sobre la disponibilidad de información que permita conocer quiénes reciben los apoyos del programa (padrón de beneficiarios). Sostiene que la información disponible cumple con una de las características solicitadas por el Coneval,⁷ pero no aclara con cuál. El documento vuelve a ser inconsistente en sus respuestas al afirmar que “en la revisión llevada a cabo no se identificaron documentos o bases de datos que permitan identificar el Padrón de Beneficiarios. Aunque podría integrarse a partir de los expedientes de las contrataciones de las obras de rehabilitación y modernización que actualmente realiza el programa en Distritos de Riego” (2013:14). Es decir, de nueva cuenta se responde afirmativamente a una de las preguntas del Coneval, pero esta respuesta afirmativa no se deriva de información del propio programa, sino de una sugerencia de mejora que en todo caso, es un resultado de la evaluación. Por tanto, la no disponibilidad de información sobre los beneficiarios debería haberse señalado también como una omisión en el diseño del programa evaluado.

- Matriz de Indicadores para Resultados (MIR)

A diferencia del análisis sobre las poblaciones del programa, en el apartado correspondiente al análisis de la MIR el informe sí diferencia claramente la valoración que se deriva de la información del propio programa, de las recomendaciones que se derivan de la evaluación. El análisis de la MIR se muestra en tres partes: desde su lógica vertical; desde su lógica horizontal; y una valoración integral.

⁷ Los TR del Coneval señalan como criterios para evaluar la información sobre los beneficiarios de los programas: *a*) que se incluyan las características de los beneficiarios establecidas en su documento normativo; *b*) que se incluya el tipo de apoyo otorgado; *c*) que esté sistematizada; y *d*) que cuente con mecanismos documentados para su depuración y actualización.

En cuanto a la lógica vertical, el informe identifica un desatino importante en la selección de indicadores usados tanto para los componentes, como para las actividades. En lo tocante a las actividades, señala que aunque es posible que casi la mitad de las acciones señaladas en la MIR cumpla con los criterios de especificación (no ambigüedad), orden cronológico y probablemente de necesidad (es decir, son acciones no prescindibles), dichas actividades están más vinculadas con la gestión y seguimiento administrativo del programa (v.g. gestión de los Oficios de Liberación de Inversiones o seguimiento al ejercicio de los programas), que al logro de objetivos del mismo, pues no son acciones vinculadas con la productividad de la tierra. Ante esta situación, el informe recomienda redefinir los niveles de Actividad y Componente de la MIR, de forma que los indicadores incluidos sean más congruentes con el objetivo del programa. En los componentes el análisis es similar, pues el informe indica que no están vinculados a la medición de la mejoría en la productividad de la tierra, sino a servicios producidos por el programa (v.g. obras de infraestructura conservada y/o modernizada en los DR).

Por último, el informe repite su sentencia cuando habla del propósito o fin de la MIR⁸ subrayando que el mismo no necesariamente conduce a una mejoría en la productividad de la tierra, al tiempo que ésta no necesariamente conduce a reducir la sobreexplotación de los acuíferos o la reducción de los volúmenes “sobreconcesionados”. El documento es claro al afirmar que se requiere revisar la MIR desde el planteamiento del Fin, el cual debiera estar encaminado más hacia la contribución del programa al uso eficiente y sustentable del agua en el sector agrícola.

Con respecto a la correspondencia entre el documento normativo del programa con el resumen narrativo de la MIR (fin, propósito, componentes y actividades), el informe enfatiza nuevamente la falta del primero, conteniendo tanto sus lineamientos como su manual de operaciones y procedimientos para regular su operación. En este sentido, no es posible identificar el planteamiento de los objetivos en ningún otro documento que no sea la propia matriz.

⁸ Esto es: “La infraestructura hidroagrícola en los Distritos de Riego es conservada y modernizada, en beneficio de los usuarios de los Distritos de Riego”.

Las preguntas sobre si estaban presentes en la MIR los criterios para evaluar la lógica horizontal del programa (i.e. medición del desempeño del programa; fichas técnicas completas; metas claras y medibles; medios de verificación, entre otros) recibieron todas respuestas afirmativas, con base en un amplio número de observaciones realizadas durante la evaluación.

Valoración de la evaluación

Que una institución con una tradición de información difusa, como la Conagua, muestre actitudes cooperativas ante un ejercicio de evaluación de uno de sus programas, es sin duda un avance en términos no sólo de rendición de cuentas, sino de ampliar la capacidad de respuesta de sus acciones, al abrir una ventana de oportunidad para la incorporación de nuevos elementos que mejoren su diseño.

Por otro lado, conviene apuntar que tales sugerencias de mejora constituyen el aporte más fuerte y novedoso de la evaluación de diseño metaevaluada, al aportar no sólo una metodología robusta para definir aspectos tales como la identificación y cuantificación de la población potencial y objetivo del programa, sino la justificación conceptual para formular nuevos indicadores en la MIR. Sin embargo, resulta necesario tomar en cuenta que la forma en la que estas recomendaciones se introducen a lo largo del informe puede ocasionar confusión, pues a veces se utilizan como justificación para responder por los contenidos del programa evaluado, cuando en realidad son alternativas de mejora.

En resumen, se considera que el informe cumple con el objetivo de brindar información metodológica y de carácter conceptual, que puede traducirse en un rediseño para la mejora del programa. Por otro lado, se considera que la decisión del Coneval de formular unos TR modelo, orientados a establecer instrumentos homogéneos, en ciertos casos genera una esquematización, sobre simplificación y redundancia de contenidos en los informes finales. Esto dificulta la emisión de juicios valorativos sobre la calidad de las evaluaciones *per se*, limitando el alcance de las metaevaluaciones a una valoración indirecta sobre dichos procesos, con base en informes esquemáticos y poco flexibles.

Conclusiones y recomendaciones parciales

Las tres evaluaciones revisadas en este apartado presentan características muy diferentes entre sí. En cambio, coinciden en que los programas a los que se aplicaron son de reciente creación (Promanp en 2012 y Estabilización de Cuencas y Acuíferos en 2013), o lo eran al momento de ser evaluados (PNANP). En consecuencia, se justifican sus evaluaciones de diseño y de indicadores, aunque en el caso del PNANP se esperaba una evaluación efectiva de las ANP consideradas en el programa.

Como se ha referido en varias oportunidades, las evaluaciones de diseño cuentan con un modelo general de TR preparado por el Coneval que fue aplicado en los dos casos de este tipo que fueron considerados; en cambio, la evaluación del PNANP no contó con ellos y sólo se acercó a las llamadas evaluaciones de indicadores en el marco de los Lineamientos del Coneval, que no tuvo participación en la misma.

En cuanto a la calidad de los equipos evaluadores, vuelve a dificultarse la evaluación ya que: *a*) en un caso no existieron evaluadores como tales y sólo se mencionan los participantes en dos reuniones internas de trabajo; *b*) en otro, fue contratado un solo especialista independiente, con formación suficiente y experiencia en la Semarnat, aunque no en el área de evaluación, y *c*) el tercero también es un experto individual del que se desconocen más datos. Los dos casos de evaluaciones externas de diseño fueron contratados por asignación directa y los montos pagados fueron de \$102 080.00 en el Promanp y \$92 220.00 en Estabilización de Cuencas y Acuíferos, a pesar de que este último trabajo fue notoriamente más complejo que el primero. La forma en que se decide la asignación y se definen los montos permanece en la opacidad.

La metodología utilizada por el PNANP se desconoce totalmente, tanto en torno al trabajo de las reuniones como a su preparación previa y elaboración posterior del informe y sus conclusiones. En los casos de diseño, uno se limitó al análisis de documentos institucionales que es lo solicitado por los TR, pero el otro refiere como principal fuente de información un cuestionario (con las preguntas de los TR) aplicado a informantes-clave de la Conagua, que no identifica ni cuantifica. Este segundo caso se valora positivamente por complementar la información documental.

La extensión efectiva de los informes (sin considerar anexos) es pequeña (varía entre 33 y 38 páginas). Éstos se presentan adecuadamente estructurados (los dos de diseño se exponen en el orden de las preguntas de los TR), bien redactados y sólo muestra ambigüedades el referido a Estabilización de Cuencas y Acuíferos que, por otro lado, es el más completo. Este último y el de Promanp se encuentran disponibles en la página web del Coneval, mientras que el PNANP aparece en el portal de la Conanp. Sólo en el caso de Promanp se localizó un documento de posicionamiento institucional como reacción ante el trabajo realizado y aceptando realizar las cinco correcciones propuestas, pero dos de éstas no fueron transformadas en ASM (Coneval, 2012c). Se desconoce si en Estabilización de Cuencas y Acuíferos se formularon aspectos susceptibles de mejora.

Se trata de trabajos de calidad muy variada: el de PNANP no evalúa realmente las ANP y sus indicadores no recuperan los tres objetivos vinculados con la sustentabilidad; el de Promanp cumple formalmente con la solicitud pero el programa es tan simple (un solo objetivo y componente) que el evaluador cae en reiteradas repeticiones; en cambio, el de Estabilización de Cuencas y Acuíferos evalúa realmente todos los aspectos solicitados por los TR en forma aceptable.

Como consecuencia de lo anterior, el sector participa de algunos problemas comunes a los ya analizados, por lo que sólo se hacen propuestas para los problemas específicos identificados: *a*) en el caso de Estabilización de Cuencas y Acuíferos se limitan a mejorar la redacción del informe para evitar confusiones y contradicciones y flexibilizar los TR; *b*) en el de Promanp, reflexionar si se debe hacer un programa para facilitar el cumplimiento de una obligación legal, remunerar la evaluación en función de su sencillez y exigir un mayor tratamiento de los problemas inmersos en el programa; y *c*) en el de PNANP mejorar la composición del sistema de indicadores para recuperar equilibradamente los tres ejes de la sustentabilidad y aplicar evaluaciones externas en las ANP con base en manuales internacionales ya probados (i.e. Pomeroy *et al.*, 2004) y adaptados al contexto y los objetivos nacionales (Rodríguez, 2014).

CAPÍTULO XVIII
Metaevaluaciones en materia de salud federal

Myriam Cardozo Brum

**Consistencia y resultados del Sistema de Protección Social en Salud.
Seguro Popular de Salud 2007**

Contexto de la evaluación

El informe de evaluación menciona que se realiza la ECR del Programa Salud para Todos “Seguro Popular de Salud”, a cargo de la Comisión Nacional de Protección Social en Salud (CNPSS) de la SS en el periodo comprendido entre el 17 de agosto de 2007 y el 24 de marzo de 2008, sin explicar la razón del mismo. Inmediatamente agrega que evaluará la consistencia del Sistema de Protección Social en Salud (SPSS), lo que genera confusión ya que el documento se refiere indistintamente al SPSS o al Seguro Popular, sin aclarar sus relaciones.

El SPSS entró en vigor en 2004 con el objetivo de que las familias sin seguridad social en salud tengan acceso a las acciones de protección establecidas por la Ley General de Salud, evitando así gastos catastróficos. Opera en las 32 entidades federativas y la incorporación al mismo es voluntaria, a solicitud del jefe de familia, con una duración de un año renovable. Su financiamiento es tripartita: gobierno federal, gobiernos estatales y aportaciones de las familias incorporadas al Sistema. En 2007 su presupuesto ejercido fue de 26 290.4 millones de pesos.

Desde su inicio contó con una amplia estrategia de evaluación, con actividades internas y externas, en materia contable, de procesos y de efectos, cuya aplicación

quedó a cargo de la Dirección General de Evaluación del Desempeño (SS, 2005). Dichas actividades fueron evaluadas en 2005 y 2006 con base en un diseño de la Universidad de Harvard, el INSP y el CIDE. Adicionalmente, en 2002 contó con una evaluación de la gestión y de impacto mediante diseño experimental realizada por el INSP; y en 2003 la empresa Berumen y Asociados realizó una auditoría técnica a su padrón. Finalmente, y en forma paralela a la presente ECR, al programa se le aplicó una evaluación complementaria en 2007.

Marco normativo

Como ya se mencionó, los Lineamientos prevén este tipo de evaluaciones y Coneval elaboró un modelo de TR consistente en 100 preguntas referentes a su diseño, planeación estratégica, cobertura y focalización, operación, percepción de la población objetivo y resultados basados en el seguimiento de indicadores, a responder mediante trabajo de gabinete¹ (Coneval, 2007b).

Equipo evaluador y recursos económicos

El equipo de evaluación estuvo coordinado por un médico por la UNAM, con maestría en salud pública de la Universidad de Texas, de larga trayectoria en el sector salud y en organismos internacionales, e integrado por dos investigadores asociados, actuarios por la UNAM (uno con maestría en salud pública por la Universidad de Boston y otro con especialidad en encuestas por muestreo del Instituto de Estadística de España, en metodología cuantitativa aplicada a la epidemiología en la UAM y en programas de mercadeo social por el Centro para el Desarrollo Social de Chicago), y dos investigadores asistentes, uno sociólogo y otro médico. El nivel académico y la formación interdisciplinaria del equipo se consideran apropiados; pero no se dispone de información en cuanto a su formación o experiencia previa en evaluación.

¹ En 2011 redujo a 51 el número de preguntas (Coneval, 2011h).

El Coneval (s/f:3) informa que se aplicó el mecanismo de invitación a cuando menos tres instancias para la asignación del contrato que recayó en la sociedad civil denominada Investigación en Salud y Demografía y el monto correspondiente fue de \$569 931.00.

Metodología de evaluación

El proceso metodológico de la investigación incluyó sólo trabajo de gabinete para el análisis de información documental; se utilizaron técnicas como la Matriz de Marco Lógico (MML) y el esquema FODA. El proceso fue guiado por las 100 preguntas incluidas en el modelo de TR correspondiente. Destaca como una actividad deseable y poco usual la realización de seis reuniones de coordinación y revisión de avances entre el Coneval, la CNPSS y el equipo evaluador, y cuatro más de presentación de resultados y con los consejeros del Consejo Nacional de Evaluación de la Política de Desarrollo Social.

Informe final

El informe final consta de 154 páginas que exponen las características generales del SPSS y los documentos normativos en que se apoya, para continuar con pleno apego a los apartados y preguntas señalados en los TR. Se destaca que todas las preguntas cuentan con una argumentación rigurosa de su respuesta y recomendaciones, incluso para mejorar las que presentan un desempeño positivo. Dispuso también de amplia documentación, aunque concentrada en trabajos institucionales.

El informe valora el diseño como adecuado pero no bien recuperado en la MIR, especialmente en torno a la definición y congruencia de su fin y propósito y de su lógica vertical y horizontal, lo que se fundamenta con detalle, se propone considerar la posibilidad de realizar una evaluación específica de indicadores y se recupera adecuadamente en las conclusiones. Las únicas debilidades que podrían señalarse en este apartado se relacionan con la limitada justificación de que el programa pueda alcanzar su propósito basada sólo en un artículo y una evaluación nacional interna (ninguna internacional), y la inexistencia de

un diagnóstico formal del problema encarado por el programa que se trata de justificar con otros trabajos de investigación y evaluaciones internas. Estos dos problemas, así como la falta de un estudio sistemático de complementariedades y no duplicación con otros programas, se señalan pero no se les da suficiente trascendencia como para incluirlos en las conclusiones.

En torno a la planeación estratégica identifica varias debilidades (falta de planes, estrategias, metas, indicadores, plazos, etcétera) y, muy especialmente, la falta de acciones y seguimiento a las recomendaciones de los últimos dos años que permitan mejorar el desempeño del programa y la vinculación con la asignación de recursos presupuestales. Se señalan algunos elementos de planeación incluidos en diversos documentos como leyes y programas, pero también se ubican inadecuadamente los Informe semestrales de resultados.

En materia de cobertura y focalización, operación, y percepción de la población objetivo, la evaluación es positiva pero se refuerza con algunas recomendaciones de mejora. En cambio, se omite recuperar algunos de los hallazgos previos en las conclusiones por temas, como la ausencia de una estrategia de cobertura a corto, mediano y largo plazo y la falta de indicadores de eficacia, eficiencia, economía, costos de operación y costo/efectividad, los que reaparecen en la conclusión global. Se menciona una mejora regulatoria de la operación pero no se la vincula con evaluaciones previas.

En relación con los resultados del programa, el documento se refiere a la información regularmente recogida y las evaluaciones externas realizadas con metodologías valoradas como rigurosas (incluyen estudios experimentales con selección por pareamiento), que muestran resultados positivos en materia de logro de objetivos pero sin metas de referencia a nivel de fin y propósito. Los resultados de la evaluación analizada se muestran con mucha claridad en un cuadro FODA que incluye la pregunta fuente de cada recomendación y la dependencia o entidad responsable de aplicarla. Éstas también se resumen, por temas, en el apartado de conclusiones, el que acaba con una conclusión general: el necesario establecimiento de un modelo integral de monitoreo y evaluación en que se resuelvan los problemas identificados.

Valoración de la evaluación

Se trata de una evaluación que cubre la totalidad de los temas solicitados por los TR y con un muy buen nivel de calidad, a la que sólo se le han identificado algunas debilidades mínimas. Este resultado es aún más positivo si tomamos en cuenta que los recursos con que contó para realizar su trabajo fueron bastante menores a los dispuestos por varias de las evaluaciones analizadas en esta parte del libro. Permite ejemplificar también la utilidad de las ECR, consideradas las más completas después de las de impacto, cuando éste se explica además de medirse.

Complementaria del Sistema Integral de Calidad en Salud 2008-2009

Contexto de la evaluación

El Programa de Acción Específico del Sistema Integral de Calidad en Salud (Sicalidad) inicia en 2007 en el marco de la planeación nacional del nuevo sexenio que reconoce el rezago en materia de calidad de servicios de salud y tiene cobertura en las 32 entidades federativas del país. El programa tiene como antecedente la Cruzada Nacional por la Calidad de los Servicios de Salud (Ruelas, citado en UAM-Xochimilco, 2009) y diversas sugerencias de la Organización Panamericana de la Salud (OPS).

En consecuencia, busca diagnosticar carencias e insuficiencias en la calidad de los servicios de salud, y formular propuestas y acciones tanto para mantener los avances logrados como para favorecer un programa de mejora continua. Dirige sus objetivos a las unidades de atención primaria y hospitales que desarrollen proyectos para mejorar la calidad, a los que otorga financiamiento o cofinanciamiento y apoya su difusión, capacitación y asociación.

A la fecha de la presente evaluación el programa contaba con una evaluación de diseño de 2007 y, en paralelo, se le aplicó durante 2008 una EED. La evaluación complementaria que se evalúa brinda muy poca información para comprender los objetivos, estrategias y acciones del programa evaluado. En general, la metaevaluación enfrentó la dificultad de localizar la información necesaria para su análisis debido a que normalmente el informe completo está disponible en la página del Coneval, pero en este caso sólo incluye la síntesis

ejecutiva del mismo, anexos e instrumentos de recolección de información. Un documento de la Dirección General de Calidad y Educación en Salud de la Secretaría de Salud aportó nuevos datos sobre la metodología utilizada, y por último, se accedió al informe completo en el portal de salud.²

Marco normativo

Puesto que, según los Lineamientos, esta evaluación se define opcionalmente en función de las necesidades e intereses de las dependencias y entidades para conocer su desempeño y mejorar su gestión, sólo es posible enfocarse a la amplia diversidad de temas que este tipo de evaluaciones ha cubierto a la fecha: evaluaciones iniciales, de diagnóstico, de diseño, de desempeño, de procesos y resultados intermedios, de indicadores, del sistema de monitoreo y evaluación, de seguimiento de metas, de cobertura y satisfacción de usuarios, de percepción de beneficiarios, de seguimiento de beneficiarios, de coordinación, de operación, técnico-financiera y de procesos, de resultados, de consecución de metas, de efectos atribuibles, metaevaluaciones y algunas temáticas como educación, reforestación, salud, nutrición, etcétera. Sorprende que algunos de estos temas corresponden a otros tipos de evaluación identificados en los Lineamientos como las de procesos, indicadores o efectos atribuibles o impacto; otras como las de diseño se volverían luego un tipo independiente.

Dada esta diversidad, el guión de la metaevaluación tendría que tomar en cuenta la solicitud concreta expresada en cada caso en los TR correspondientes, pero éstos no están disponibles en el caso de Sicalidad, por lo que sólo es posible analizar lo que el equipo evaluador efectivamente incluyó en el informe correspondiente, que afirma se apegó plenamente a lo solicitado. El caso de

² La síntesis ejecutiva está en la página del Coneval [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>]. El documento metodológico se agrega en la página de la Dirección General de Calidad y Educación en Salud de la Secretaría de Salud [http://www.calidad.salud.gob.mx/site/calidad/dmp-process_07H.html]. El informe completo (63 páginas de extensión) se ubicó en el portal de salud [http://portal.salud.gob.mx/codigos/columnas/evaluacion_programas/pdf/EXT09_SICALIDAD_IF.pdf], fecha de consulta: 22 de octubre de 2014.

Sicalidad fue tipificado como una evaluación técnico-financiera y de procesos; su informe final explicita los temas cubiertos: los aspectos técnico-financieros y los procesos de los componentes de: seguridad del paciente, calidad de la gestión, y proyectos de apoyo y reconocimiento financiados en 2008.

Equipo evaluador y recursos económicos

La evaluación fue realizada en la Universidad Autónoma Metropolitana, Unidad Xochimilco, pero no existe información por parte del Coneval del proceso de convocatoria, selección y contratación del equipo, su monto ni sobre la formación y experiencia de sus integrantes.

En cambio la Dirección General de Calidad y Educación en Salud de la SS informa que participaron dos doctores como responsables de cada una de las partes de la evaluación y una licenciada como coordinadora de la logística del trabajo de campo; así mismo, que la contratación fue por asignación directa y un monto de \$3 758 344.00.

Por su parte, el portal de salud agrega la participación de un cuarto maestro como integrante del equipo y de 12 asistentes de investigación con nivel de licenciatura o maestría. Los responsables de la evaluación cuentan con sólida formación en medicina, enfermería, medicina social y ciencias sociales, pero no existen referencias sobre su experiencia previa en trabajos de evaluación.

Metodología de evaluación

La síntesis ejecutiva del informe señala como metodología utilizada para el análisis técnico-financiero un proceso de diseño de matrices que sintetizan la información de los proyectos. Luego explica que la información se obtuvo de proyectos de una muestra cuya composición se señala en términos generales. La identificación concreta de las instituciones participantes aparece en el Anexo 1 y en el informe completo, mientras el tipo de muestreo aplicado (por conveniencia y con base en criterios) se explica en la nota metodológica. En la parte de análisis se refiere al otro componente de la evaluación, o sea, algunos procesos relativos a la gestión de proyectos. Se trata de un proceso que se debe

ir reconstruyendo porque no se explica en forma clara y ordenada en ninguno de los documentos revisados.

El informe presenta por separado la metodología utilizada en el estudio de los procesos. En éste la información se obtuvo mediante un trabajo de campo que recogió la observación directa y las respuestas a entrevistas y a un cuestionario a funcionarios y beneficiarios con base en una muestra, que parece ser la misma utilizada para el estudio técnico-financiero. En este caso el análisis se hizo con el programa Atlas Ti y el SPSS. Sorprende que se mencione que el diseño de estos instrumentos contó con el aval de la Dirección General de Evaluación del Desempeño de la SS, cuando es fundamental que un evaluador externo tenga libertad plena para hacerlo.

La síntesis ejecutiva no informa sobre los datos efectivamente recogidos y analizados, lo que impide valorar la calidad del proceso más allá de su planeación.³ El Anexo 1 brinda información más detallada de los proyectos pero es meramente descriptiva.

Informe final

De los tres componentes que integran el informe de 63 páginas de extensión, el de calidad sobresale como el principal por la cantidad de objetivos específicos que aglutina. En él se revisan los compromisos, la capacitación, los premios y estímulos otorgados, el papel del gestor en esta materia, y la calidad de la atención en hospitales y centros de salud. En cuanto al primer componente enunciado, sólo hay un objetivo específico y se limita a identificar las acciones que realizan las unidades hospitalarias que ejecutan el Programa de Seguridad del Paciente, lo que resulta insuficiente para fines evaluativos. Por último, el denominado tercer componente, en realidad forma parte de los objetivos de calidad, puesto que a su logro se dirigen los proyectos de apoyo y reconocimiento.

³ Además no aclara los acrónimos y siglas utilizados, lo que dificulta mucho la comprensión del trabajo. Por ejemplo, se habla mucho de proyectos de AG, pero nunca se dice qué es. En cambio esa información aparece en el informe final.

El informe pasa directamente de los objetivos y metodología a los resultados del trabajo, los que acompaña de las conclusiones y recomendaciones para cada uno de los objetivos específicos propuestos. Para el análisis técnico-financiero los resultados se presentan organizados en matrices de dos componentes (fortalezas y debilidades). Puesto que el informe no tiene un apartado en que se expliquen los contenidos del programa, de los componentes analizados, de las áreas de la SS involucradas en él y tampoco se informa de los datos recogidos, las matrices sólo constituyen un conjunto de afirmaciones cuyo fundamento generalmente no se presenta. En Anexos se puede ver que existe información más amplia, por ejemplo, para analizar el perfil de los Gestores de Calidad, pero en el informe no hay mención alguna a este tema.

Es importante recordar que una evaluación no se reduce a los resultados de encuestas y entrevistas, éstas constituyen un insumo más a analizar para que el evaluador arribe a su propia conclusión y valoración. Pero en la parte referida a procesos el informe se limita a exponer las manifestaciones de los entrevistados, sin evidenciar el compromiso del evaluador para verificarlas y valorarlas.

En cuanto a las recomendaciones, éstas son muchas y se organizan por temas; algunos ejemplos son: incrementar el presupuesto destinado para Acuerdos de Gestión (AG), promover que el financiamiento se distribuya equitativamente por tipo de gasto y agilizar la llegada oportuna de los recursos financiados, seleccionar proyectos de AG con base en un calendario por regiones del país y brindar mayor asesoría, dar prioridad en la selección a los proyectos de capacitación enfocados a calidad, formular una agenda de capacitación con tres líneas generales (programa Sicalidad, temas de calidad y capacitación técnica), cumplir los plazos fijados en las convocatorias, dar mayor difusión a los derechos del paciente, difundir entre las unidades no acreditadas los procesos exitosos de las unidades, dar mayor seguimiento a las unidades acreditadas para garantizar los estándares de calidad, establecer algún estímulo a los gestores de calidad exitosos, recuperar las buenas prácticas y socializarlas, hacer más “amigable” la página electrónica del programa, entre otros. Se trata de una variedad con distintos grados de especificidad: algunos implican sólo cumplir lo que ya está establecido, varios son muy generales y no refieren una estrategia para alcanzarlos, otros son muchos más concretos.

Valoración de la evaluación

La síntesis ejecutiva del informe no brinda elementos para comprender las conclusiones a las que arriba. Si bien en el informe completo hay más datos, la síntesis debe permitir el acceso a información menos detallada pero no a la falta de eslabones que permitan comprenderla, en tanto documento independiente.

En la parte técnico financiera se concluye que existe coherencia en el diseño técnico de los proyectos y que el presupuesto es insuficiente y poco equitativo, lo cual amerita al menos un par de cuadros con la información de base, siguiendo los TR de otros tipos de evaluación que cubren estos aspectos (diseño, ECR).

Nuevamente, en relación con la evaluación de procesos y en ausencia de los TR específicos, se espera que se analicen y valoren todos los puntos señalados en los TR de ECR y las específicas de procesos, como planeación, programación, difusión, solicitud de apoyos, selección de beneficiarios o de proyectos, distribución de apoyos, entrega de apoyos, seguimiento a beneficiarios, entre otras. Sin embargo, no es seguro que los TR los hayan solicitado, si es que existieron. Sólo en el apartado del Programa de estímulos, que fue el único en que se entrevistó a beneficiarios, se presentan algunas conclusiones sobre estos aspectos y, en el documento completo, en materia de diseño, convocatoria y selección de proyectos.

En síntesis, falta analizar varios procesos sustantivos, fundamentar juicios e imprimir al documento un carácter más analítico y evaluativo, lo que hubiera sido perfectamente posible con los recursos dispuestos.

Conclusiones y recomendaciones parciales

Las dos evaluaciones revisadas en este apartado corresponden a programas con una antigüedad de dos o tres años al momento de ser evaluados y que ya habían sido objeto de trabajos previos, lo que justifica el tipo de evaluación seleccionado, además de que las complementarias pueden aplicarse en cualquier momento que las dependencias y entidades lo requieran.

Mientras las ECR cuentan con un modelo de TR estándar, la complementaria necesita su redacción específica, pero en el caso de Sicalidad no se localizó el documento correspondiente.

Respecto de la calidad de los equipos evaluadores contaron con una adecuada formación pero ninguno de los dos ofreció evidencias sobre su experiencia en evaluación. En el caso del Seguro Popular de Salud la convocatoria fue realizada por invitación a por lo menos tres instancias y se carece de información sobre el proceso de selección; en el de Sicalidad no existe dato alguno. El monto contratado en el primero ascendió a \$569 931.00 y en el segundo a \$3 758 344.00; como es habitual, no existe justificación de los mismos.

La metodología utilizada en la ECR fue el análisis documental, en apego a los TR, pero agregó una serie de reuniones de trabajo que se valoran en forma muy positiva. En el caso de Sicalidad incluyó como fuentes la observación directa, así como entrevistas y cuestionarios, cuyo muestreo no queda totalmente aclarado en el informe, al igual que la manera en que fue analizada la información recogida.

El informe de la ECR tiene una extensión suficiente para el trabajo solicitado (154 páginas), es muy ordenado en su exposición y el informe completo está disponible en la página web del Coneval, como sucede ya habitualmente; en cambio, se esperaría un informe más extenso en el caso de una evaluación complementaria, pero el de Sicalidad sólo cuenta con 63 páginas, es un poco confuso y el Coneval sólo publicó su resumen. No se contó con un documento de respuesta institucional a las evaluaciones ni se conoce si éstas dieron lugar a ASM. En síntesis, la calidad del primero resulta muy buena, mientras el segundo hubiera requerido mejorarse.

Como consecuencia de lo anterior, más allá de algunos problemas comunes a los sectores ya analizados, a la presente ECR sólo se le recomienda una mayor recuperación en las conclusiones finales de los problemas identificados durante el proceso, y a la evaluación complementaria, una mayor cobertura de temas, extenderse más en su tratamiento, fundamentación y valoración de los hallazgos.

CAPÍTULO XIX
Metaevaluaciones de desarrollo rural federal

Osmar Medina Urzúa

**Consistencia y resultados del Programa de Desarrollo de Capacidades,
Innovación Tecnológica y Extensionismo Rural 2011-2012**

Contexto de la evaluación

El Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural (PDCITYER) fue instaurado en 2011 derivado de una reestructuración programática efectuada al interior de la Sagarpa ese mismo año, y está integrado por elementos de los antiguos programas de Soporte¹ y de Fortalecimiento a la Organización Rural.

[Su objetivo es] fortalecer las capacidades técnicas y administrativas de las unidades económicas agropecuarias, pesqueras y acuícolas para que mejoren sus procesos productivos y organizativos a través del otorgamiento de apoyos en servicios de asistencia técnica, capacitación, extensionismo; en proyectos de investigación y transferencia de tecnología y en fomentar el desarrollo gerencial de las organizaciones sociales y de los comités sistema producto (Sagarpa, 2011:42).

¹ El Programa Soporte, creado en 2008, se orientaba a mejorar las capacidades de los productores mexicanos, así como al entorno sanitario y biológico en el que los pescadores desarrollan sus actividades (Sagarpa, 2011).

En 2011 el programa tuvo una asignación presupuestal de \$5 048.69 millones y operó en concurrencia de recursos con los gobiernos de los estados (en todas las entidades federativas del país, incluyendo al Distrito Federal). Estos últimos fungen como instituciones ejecutoras, encargadas de la operación de los programas. Para cumplir su función, se ha creado una Comisión Nacional de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural y comisiones estatales, donde los gobiernos de los estados coordinan y supervisan el programa, mientras la Sagarpa, por medio de sus delegaciones, se encarga de normar y supervisar la instrumentación del Programa, como unidad responsable (FAO, 2013).

Al tratarse de un programa “nuevo”, siendo estrictos, no se pueden mencionar evaluaciones anteriores hechas al mismo; sin embargo, considerando que el PDCITYER aglutinó como componentes propios a programas (o por lo menos gran parte de éstos) que contaban con evaluaciones en años anteriores, se mencionan las evaluaciones más recientes efectuadas a dichos componentes: evaluaciones de diseño 2008 y EED 2009, hechas al Programa de Soporte al Sector Agropecuario por la Academia Mexicana de Auditoría Integral y al Desempeño, AC, y el Colegio de Postgraduados respectivamente; EED 2010 hechas al programa Soporte y al programa de Fortalecimiento a la Organización Rural, ambas realizadas por el CIESAS-Occidente; y, en menor medida, aquellas hechas a los componentes y actividades del ya mencionado Prodesca (capítulo 13) que era contemplado como parte de las evaluaciones del programa de desarrollo rural 2002-2006 de la Alianza para el Campo.

Marco normativo

En la creación del PDCITYER se retoman los mandatos de la LDRS (2001), que en sus artículos 42 y 48 establece un Sistema Nacional de Capacitación y Asistencia Técnica Rural Integral (Sinacatri) y un Servicio Nacional de Capacitación y Asistencia Técnica Rural Integral (Senacatri) recursos para que los productores contraten servicios con los componentes.

Respecto a la ECR 2011-2012 del PDCITYER, ésta se llevó a cabo con fundamento en lo establecido en los artículos 72 al 80 de la LGDS (2004); en el artículo 78 de la Ley Federal de Presupuesto y Responsabilidad Hacendaria

(2006, modificada en 2008); a lo dispuesto en el capítulo primero, título tercero, de los Lineamientos, y en el PAE, en el que se establecen las evaluaciones que se llevarán a cabo cada año, los programas a los que se aplicarán y en el que se incluyó la realización de una ECR al Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural.

Debido a que los TR genéricos diseñados por el Coneval para este tipo de trabajos (Coneval, 2011h) fueron insuficientes para proporcionar a la Sagarpa toda la información que requería, se decidió elaborar dos documentos: el primero, siguiendo el formato, preguntas y TR establecidos por el Coneval, y el segundo (que no se publicó), incluyendo la problemática, diversidad y complejidad del programa de manera integral, y respondiendo a la totalidad de preguntas de cada uno de los componentes que integran al Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural.

Equipo evaluador y recursos económicos

Para la realización de la presente evaluación, la Dirección General de Planeación y Evaluación (DGPE) de la Sagarpa no realizó ningún tipo de convocatoria sino que hizo un análisis interno sobre los posibles evaluadores externos que cumplieran con los requisitos mínimos de elegibilidad para su contratación, señalados en el capítulo III de los Lineamientos. Decidió dirigir su invitación a quien fungió posteriormente como coordinadora del trabajo, en el marco institucional del CIESAS-Occidente, con base en su experiencia en el tema y en cinco evaluaciones previas hechas a programas de la Sagarpa. La evaluación fue realizada como parte de un convenio de colaboración entre esta institución y la Sagarpa. El costo de la misma fue de \$730 000.00, según lo reportado en el informe final, y cubrió los dos documentos antes mencionados.

Respecto al equipo de evaluación, éste fue coordinado por una doctora y maestra en antropología por la Universidad de Chicago con amplia experiencia en evaluaciones a nivel nacional de programas del gobierno federal de desarrollo rural, incluyendo las EED 2010 hechas al programa Soporte y al programa de Fortalecimiento a la Organización Rural. Entre los principales colaboradores se encuentra un maestro y doctor en desarrollo rural por la UAM-Xochimilco, entidad donde también es académico del posgrado del mismo nombre; además

de dos doctorantes en ciencias sociales del CIESAS con maestría y especialidad en temáticas relacionadas a la antropología y una maestra en antropología social (antropóloga médica) por el CIESAS, con experiencia en grupos vulnerables, comunidades indígenas y sectores diversos. Si bien los últimos dos evaluadores carecen de una especialización sectorial en el ámbito agropecuario, se considera que este aspecto es compensado por la experiencia de la coordinadora de la evaluación y su principal colaborador.

Metodología y procesos de evaluación

Los evaluadores externos entrevistados manifestaron que tuvieron buena comunicación con la DGPE de la Sagarpa, lo que facilitó la entrega de información y la organización del trabajo de campo. Lo mismo sucedió con el Coneval, salvo en relación con los TR que los evaluadores y la Sagarpa consideraron demasiado acotados, mientras el Consejo prioriza la homogenización con fines comparativos y la simplificación ante la magnitud de su encargo, por lo que insistió en que la evaluación se hiciera en la forma preestablecida para todas las ECR. Esto llevó a los evaluadores a enfrentar también el problema de la dificultad de responder a algunas preguntas de los TR de manera global y dicotómica, cuando tendrían que precisarse componentes en que la respuesta es positiva y otros en que resulta negativa.

El informe final de la evaluación no menciona el enfoque metodológico utilizado para realizar la evaluación así como los propósitos de la misma; sin embargo, esto se expresa en los TR. De este modo, la ECR sigue el esquema propuesto por el Coneval para evaluar la consistencia y orientación a resultados del programa y proveer de información que retroalimente su diseño y gestión.

En consecuencia, la evaluación se llevó a cabo mediante el análisis de seis grandes áreas del programa: diseño, planeación y orientación a resultados, cobertura y focalización, operación, percepción de la población atendida y medición de resultados. Cada área es analizada respondiendo a 51 preguntas específicas preestablecidas conforme a los Lineamientos y TR emitidos por el Coneval para este tipo de evaluaciones. Todas fueron sustentadas con evidencia

documental y haciendo explícitos los principales argumentos que las fundamentan y 34 de éstas fueron respondidas mediante un esquema dicotómico (sí o no).

En el trabajo se utilizó investigación cualitativa realizada en gabinete y se aplicaron entrevistas a actores clave. En la selección de los últimos se consideraron tanto las personas propuestas por la Sagarpa a nivel central y estatal, como las sugeridas por la entidad evaluadora.² Las entrevistas fueron aplicadas en dos estados³ por cada región del país (norte, centro y sur) para representar la diversidad geográfica del mismo. Se utilizó un cuestionario semiestructurado al que no se tuvo acceso; sin embargo, dada la información que de él fue recuperada en la evaluación, se considera que fue fundamental para la justificación de las respuestas, principalmente en el apartado de operación.

Al ser un programa que apenas llevaba alrededor de siete meses operando al momento de la evaluación, la disponibilidad de insumos (las MIR, reglamentos, planes estratégicos, manuales de operación) para realizarla era limitada y heterogénea. El equipo evaluador enfrentó esta problemática analizando paulatinamente la información que iba recibiendo. A esta dificultad se agregó la desconfianza por razones políticas y el desconocimiento de los operadores a nivel estatal sobre los propósitos de la evaluación y sus mandatarios, así como cierta resistencia por estar decidida a nivel central, en lugar de responder a una necesidad propia. Dados los tiempos limitados con que se contó para realizar la ECR, se juzga que fue prudente el ejercicio realizado ya que no era posible recopilar *a priori* toda la información necesaria para dar respuesta a las 51 preguntas establecidas en los términos de referencia.

Por otro lado, ante un programa constituido por un conjunto de acciones (componentes) y recursos orientados a solucionar distintas problemáticas de un mismo ámbito pero de poblaciones diferentes, el equipo evaluador decidió establecer un criterio para medir y valorar el nivel global de cumplimiento del programa mediante un primer párrafo que expresara si el programa (enten-

² Información obtenida en las entrevistas aplicadas a los evaluadores externos del programa dado que en la evaluación no se hace explícito cómo se eligieron las personas a entrevistar.

³ En una de las entrevistas se mencionaron Sonora, Jalisco, Chiapas y Veracruz como entidades que fueron visitadas para aplicar las entrevistas a nivel operativo pero el entrevistado no recordó los estados faltantes y éstos no aparecen en el informe.

dido como un conjunto de componentes) cumplía con lo estipulado en el conjunto de preguntas dispuestas en los TR, sin entrar al análisis detallado por componentes. El segundo criterio fue complementar la información anterior con los componentes concretos del programa que cumplieran lo expresado en cada pregunta. El criterio anterior permitió, por un lado, cumplir con lo solicitado en los TR y, por otro, generar información a nivel de componentes, de mayor utilidad inmediata para la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.

En conjunto, se considera que el plan de trabajo instrumentado, los métodos de recolección y tratamiento de datos fueron adecuados y apegados a lo establecido en los TR, salvo en lo referente a la intromisión de la Sagarpa en la selección de informantes-clave.

Informe final

El informe de evaluación está compuesto por 174 páginas (contando sus respectivos anexos) donde se da respuesta a la totalidad de preguntas establecidas en los TR; justificando cada una de éstas con base a los documentos descritos en la bibliografía, los anexos y la información recabada por la investigación cualitativa realizada.

En materia de diseño se analiza correctamente su objetivo general, la consistencia interna de los objetivos específicos de los componentes del programa y su congruencia externa con los objetivos nacionales y sectoriales. Se encuentra que el problema central que el programa pretende atender no está claramente definido, existiendo la necesidad de elaborar un nuevo⁴ diagnóstico que lo precise, así como sus causas y consecuencias, para lo que se sugiere emplear un diagrama de árbol. Lo anterior, se pudo complementar abundando sobre la débil lógica vertical identificada en la MIR del programa.

Otro problema señalado en el informe es que el programa no tiene cuantificada su población potencial y población objetivo, lo que a su vez provoca que no se tengan mecanismos claros de focalización.

⁴ La evaluación infiere que el Programa parte del Diagnóstico Preliminar de la FAO, 2010 y del análisis del extensionismo agrícola en México de la OCDE 2011 (Sagarpa, 2012).

Además, el equipo identificó correctamente las inconsistencias presentes en la depuración y actualización del Sistema Único de Registro de Información (SURI) de la Sagarpa, las que fueron detectadas al aplicar las entrevistas a los operadores del programa, quienes explicaron que éstas eran provocadas por problemas de comunicación del nivel central de la Sagarpa con las áreas operativas.

En lo que se refiere a la planeación estratégica se señalan las carencias de un manual de procedimientos específico que integre el conjunto de objetivos y apoyos de sus componentes y de un plan anual integral entre los anteriores y sus distintos apoyos en torno a una misma cadena productiva, según las características de cada región y los actores involucrados en ella.

En términos de la operación, también la aplicación de entrevistas a los funcionarios de nivel operativo permitió detectar problemas en la transferencia de recursos a las instancias ejecutoras y/o a los beneficiarios y la falta de un diagrama de flujo del proceso general del programa (aunque algunos de sus componentes sí lo tengan).

Respecto a la percepción de la población, el informe sólo menciona que esa información no existe; en cuanto al seguimiento de las recomendaciones de evaluaciones externas hechas al programa Soporte, señala brevemente que se ha cumplido con los ASM; y en materia de resultados del programa consideró que su medición era limitada.

Las conclusiones recogen las debilidades encontradas a lo largo de la evaluación y sugieren recomendaciones para cada tema. Esto es complementado con el análisis incluido en el Anexo 17 “Principales fortalezas, oportunidades, debilidades, amenazas y recomendaciones”. El informe muestra de manera correcta que los principales problemas del programa provienen de su origen debido a las debilidades señaladas en materia de diseño: identificación del problema central, indefinición y falta de cuantificación de sus poblaciones, ambigua cobertura y focalización y limitada medición de resultados obtenidos por el programa.

Valoración de la evaluación

Se sugiere continuar precisando los criterios para dar respuesta a las preguntas de evaluación de los TR porque su ambigüedad permite que ante la misma situación

dos equipos de evaluadores respondan en forma opuesta.⁵ También flexibilizar y ampliar el contenido de los TR y permitir que la evaluación se haga a nivel de componentes (no del conjunto del programa), como proponen la Sagarpa y los evaluadores, apoyados en que se hacían de esa manera en el periodo 2001-2006.

Dadas las dificultades señaladas para la realización del trabajo, se reconoce el esfuerzo realizado por el equipo evaluador para obtener información de utilidad y presentarla de manera clara, pero se señalan algunas debilidades de su informe:

- a) No menciona el enfoque metodológico utilizado para realizar la evaluación, sus propósitos, y los criterios utilizados para seleccionar a los actores, Lo anterior se debe en parte al limitado espacio de respuesta para cada pregunta establecido en los TR que determina la estructura de la evaluación, lo que se trató de superar con el mencionado informe paralelo que analiza más ampliamente el programa.⁶
- b) Tampoco explicita los criterios y métodos utilizados para seleccionar a los actores clave, y no adjunta los instrumentos diseñados para la recolección de datos (i.e. guiones de entrevista).
- c) Todos los componentes fueron analizados en igualdad de condiciones siendo que algunos tienen mayor peso en el conjunto (i.e. transferencias directas a los productores), por lo que se considera que su tratamiento debió tomarlo en cuenta.

Adicionalmente, se considera necesario mejorar la transparencia, publicación y divulgación de los productos derivados de las evaluaciones como el documento complementario realizado de manera paralela a esta evaluación o la posición institucional de la Sagarpa ante el informe, ya que éstos no se encuentran publicados en las páginas de Sagarpa ni del Coneval;⁷ y por lo tanto, no se tiene acceso a toda la información generada por la evaluación.

⁵ Por ejemplo, como ya se mencionó, el PDICITYER tiene antecedentes bajo otros nombres, lo que permitió a sus evaluadores afirmar que utiliza los informes de evaluaciones externas previas; en cambio, el programa que se revisará a continuación, ante idéntica situación, fue considerado nuevo y se respondió que la pregunta no era aplicable.

⁶ El contenido señalado fue mencionado por uno de los entrevistados ya que no se tuvo acceso a dicho documento.

⁷ Consultas efectuadas el 10 de mayo de 2014.

Consistencia y resultados del Programa de Apoyo al Ingreso Agropecuario: Procampo para Vivir Mejor 2011-2012

Contexto de la evaluación

El Programa de Apoyo al Ingreso Agropecuario (Procampo) Para Vivir Mejor⁸ remonta sus antecedentes a fines de 1993, por lo que se constituye en uno de los programas más longevos de la Sagarpa y es una de las “principales herramientas de política pública sectorial y por sus asignaciones presupuestales el Programa Federal con mayor población rural atendida” (Sagarpa, 2014:1). El programa surge como un mecanismo de transferencias directas al ingreso de los agricultores mexicanos al tener que competir con productores extranjeros que reciben subsidios de sus respectivos gobiernos en el marco de la firma del Tratado de Libre Comercio de América del Norte (TLCAN) (Coneval, 2013f).

En 2011 el programa fue reestructurado y se le adicionaron componentes que anteriormente operaban de manera separada; por lo cual, fue considerado como un programa heterogéneo y “no es posible considerarlo como una unidad” (Coneval, 2013f:10) dado que atendía cuestiones como el ingreso, apoyos para diésel agropecuario-marino (además de gasolina ribereña) y la capitalización mediante la adquisición de maquinaria y equipo agropecuario. Además, cada componente del programa tenía distinta forma de operación así como diferentes unidades responsables, instancias ejecutoras y cobertura territorial.

Procampo contó en 2012 con un presupuesto de \$17 115.39 millones; monto que significaba la mayor cantidad de recursos presupuestada originalmente⁹ para los programas de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.

Al igual que el PDCITYER, al ser un programa reestructurado y por lo tanto “nuevo”, siendo rigurosos, no se podría hablar de evaluaciones anteriores

⁸ En la actualidad, componente del Programa de Fomento a la Agricultura Proagro Productivo.

⁹ Posteriormente, con las modificaciones hechas al presupuesto, resultó ser el segundo con mayor presupuesto ejercido, superado por el Programa de Prevención y Manejo de Riesgos 2011.

hechas al mismo; sin embargo, considerando que el Procampo ha mantenido a las transferencias directas al ingreso de los productores como su componente principal, se considera que ha sido evaluado constantemente. En consecuencia, el programa cuenta con: ECR para los años 2007 y 2011, EED en 2008, 2009, 2010 y 2012; una evaluación estratégica en 2009 y otras como parte del grupo de programas de Fomento Agrícola de la Alianza para el Campo 2001-2006. No está de más mencionar que a parte de las evaluaciones oficiales, el Procampo ha sido sujeto desde su creación a diversos¹⁰ estudios por la academia que han intentado evaluar sus resultados e impactos.

Marco normativo

Es el mismo presentado en la evaluación anterior.

Equipo evaluador y recursos económicos

La evaluación fue realizada por el Proyecto de Evaluación y Análisis de Políticas de la FAO-México como parte del Convenio de Cooperación Técnica que sostiene dicho organismo con la Sagarpa. El costo de la misma fue de \$347 724.77, según lo reportado en el informe final de la evaluación. Cabe destacar que éste es aproximadamente la mitad del precio pagado por otras evaluaciones del mismo tipo y en el mismo año¹¹ debido a que el convenio de colaboración antes mencionado permite ahorrar recursos al prorratear costos (i.e. personal administrativo, uso de las oficinas, personal directivo, etcétera) entre diversos estudios.

En relación con el equipo de evaluación, éste presenta tanto conocimiento técnico respecto de los instrumentos y metodología de evaluación como conocimiento especializado en el ámbito rural. La coordinación del mismo estuvo a

¹⁰ Véanse Schwentesius *et al.* (2007), Merino (2009) y Taylor y Yúnez (2013).

¹¹ El costo de la ECR 2011-2012 del PDCITYER fue de \$730 000.00, y el del programa de Prevención y Manejo de Riesgos de \$704 735.00.

cargo de un doctor en economía por la Universidad de Minnesota, con maestría en economía agrícola por el Colegio de Posgraduados (Colpos) ; el cual cuenta con amplia experiencia en la evaluación de la política sectorial, siendo director del proyecto FAO-Evaluación y consultor del Banco Interamericano de Desarrollo.

El equipo integró también a otros dos consultores de FAO: uno de ellos maestro en economía por el CIDE, con licenciatura en economía agrícola por la Universidad Autónoma Chapingo, que ha participado en diversos estudios y análisis del sector agropecuario, y la segunda, una licenciada en economía por el ITAM. Ambos con experiencia en evaluaciones hechas a los programas de la Sagarpa. En consecuencia, el equipo se considera apropiado.

Metodología de evaluación

Tanto el objetivo de la evaluación como los TR fueron los mismos mencionados en el caso anterior, por lo que suscitaron similar diferencia entre la Sagarpa y los evaluadores con el Coneval. También se reitera la buena comunicación con la DGPE de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.

Para responder a las preguntas de los TR, los principales métodos utilizados en la evaluación analizada fueron el trabajo de gabinete apoyado en información proporcionada por el programa, la dependencia o la entidad y la aplicación de entrevistas a actores-clave en cuatro estados de la República Mexicana para la obtención de información cualitativa que permitió dar sustento a las preguntas relativas al ámbito de operación. Además, a diferencia del trabajo anterior, los evaluadores elaboraron un algoritmo que ponderara el cumplimiento de cada componente con su peso relativo, determinado en función de su presupuesto.¹²

Durante la evaluación se enfrentaron problemas que afectaron su calidad, como la heterogeneidad de la estructura del Procampo. El programa operaba como un todo globalizador integrado por distintos componentes, cada uno con una lógica, población y problemática distinta. Además, no tenía una unidad global responsable, sólo contaba con unidades ejecutoras distintas para cada uno

¹² Se respondió afirmativamente sobre el conjunto cuando 60% de las repuestas individuales ponderadas cumplía el ítem y “no” en el caso contrario.

de los componentes. Otra restricción fue la carencia de información suficiente dada la reciente creación del programa, además de tener que responder dentro de un espacio limitado.

Por último, los resultados de la evaluación fueron presentados solamente ante el área de evaluación de la Sagarpa y los encargados del programa, ya que los evaluadores no fueron convocados para hacerlo ante otras instancias; también fueron entregados como estipula la Ley General de Desarrollo Social.

Informe final

La evaluación presenta una introducción clara de los objetivos, métodos y marco normativo de evaluación, así como un resumen del programa y de los hallazgos. Posteriormente responde, con respaldo en evidencia, la totalidad de las preguntas y secciones que componen el instrumento de evaluación a lo largo de 182 páginas.

El énfasis de la evaluación recayó en la parte de diseño, que fue abordada ampliamente, lo que se considera adecuado dadas las debilidades centrales del programa. El informe señaló enfáticamente la limitada lógica vertical del mismo, analizada mediante la MML,¹³ y sus consecuencias en la definición ambigua de la población objetivo para el programa como un todo, la heterogeneidad entre las poblaciones objetivo de cada uno de los componentes que integran el programa y la ausencia de una metodología de focalización. En consecuencia se recomendó discontinuar los componentes que no son justificables teórica ni empíricamente así como reestructurar el programa a partir de la definición (inferida en la evaluación) del problema que se busca atender.

Con respecto a la planeación, se indica la carencia y necesidad de un plan estratégico de mediano y largo plazo, en el que se establezcan metas, tiempos y acciones para el logro del Fin y Propósito del programa. Se considera que hubiese sido apropiado ligar nuevamente esta recomendación a la dirigida a su reestructuración, incluyendo el establecimiento de una unidad responsable para todo el programa (aspecto que sí se menciona en el informe), que sería la encargada de generar dicho plan.

¹³ El informe no explicita que se utilice la MML, esta aseveración es producto del análisis de la información dada como argumento para responder las preguntas de diseño de la evaluación.

En torno a la cobertura y focalización, se parte de reiterar las debilidades mencionadas anteriormente en el tema de diseño. A partir de esto, se sugiere la aplicación de criterios de focalización, tomando en consideración las características socioeconómicas de la población potencial, que guíen la aplicación de los recursos.

En materia de operación la investigación cualitativa instrumentada permitió evidenciar que, si bien existe información¹⁴ sistematizada del programa, ésta no es compartida entre los componentes, que generan sus propias herramientas informáticas. Se considera que este problema se debió enfatizar más porque la falta de información socioeconómica de los solicitantes provoca que no sea posible conocer si éstos realmente presentan el problema que se busca resolver ni medir los impactos del programa en sus beneficiarios; aspecto fundamental dada su antigüedad y el monto de sus recursos.

Respecto de la percepción de la población, se menciona la carencia de instrumentos para medir el grado de satisfacción de los beneficiarios; en consecuencia, se considera adecuada pero incompleta la propuesta de evaluarla porque no indica la forma de hacerlo. Por ejemplo, se pudo sugerir la aplicación de una encuesta que midiera el nivel de satisfacción de los beneficiarios en términos de oportunidad (tiempo y forma) después de la entrega de apoyos, así como su eficacia para solucionar los problemas que originaron la solicitud, la atención recibida, entre otros.

Posteriormente, este aspecto fue una de las dos recomendaciones retomadas por la Sagarpa para transformarse en aspectos susceptibles de mejora.

En relación con la medición de resultados, el análisis es limitado, repetitivo, y el de menor contenido de toda la evaluación. Las respuestas dadas a las preguntas de esta sección carecen de profundidad, una misma respuesta se repite para varias preguntas y no se proporciona información nueva para la Sagarpa. El informe se limita a señalar que el programa “documenta sus resultados a nivel de Fin y de Propósito mediante la MIR, pero aún no es posible valorar el avance de dichos objetivos, ya que la temporalidad de las mediciones es trianual” (Sagarpa, 2012b:160). Sin embargo, el tema pudo haber sido explorado de manera más amplia, destacando las dificultades encontradas para realizar evaluaciones de

¹⁴ Sólo el SURI tiene información en común.

impacto al programa a pesar del tiempo que lleva en operación y vinculándolas con las recomendaciones que se hacen respecto a “contar con información socioeconómica de población que no es apoyada por el Programa” (Sagarpa, 2012b:151) y la precisión de la población objetivo.

Se considera que los hallazgos y conclusiones de la evaluación muestran, de manera resumida en un esquema FODA, los aspectos centrales encontrados a lo largo de la evaluación y permiten una revisión ejecutiva (rápida y fácil) de la misma, aspecto positivo para su uso en la toma de decisiones.

Valoración de la evaluación

Se concluye que los resultados de la evaluación son consistentes con las recomendaciones emitidas y sirvieron para realizar precisiones en la definición de la población objetivo, la desaparición de componentes valorados de manera negativa por sus efectos regresivos, así como en el otorgamiento de subsidios diferenciados con base en las características socioeconómicas de los productores (Sagarpa, 2012b).

En consecuencia, se reconoce el esfuerzo de esta evaluación por dar cabal cumplimiento a los TR del Coneval en un programa heterogéneo y por proporcionar mayor información de utilidad para los funcionarios de la Sagarpa, por lo que se reitera la conveniencia de flexibilizar y ampliar el contenido de los términos de referencia.

Sin embargo, al igual que se recomendó en la evaluación anterior, es importante que todos los documentos relacionados con la misma sean accesibles por internet y se reitera la necesidad de continuar precisando los criterios para responder a las preguntas de los TR por la situación planteada como recomendación al PDCITYER: mientras la evaluación del Procampo considera que su reestructuración lo hace nuevo y por lo tanto no son aplicables preguntas sobre sus antecedentes, el PDCITYER, en idéntica situación, responde con base en la información de los viejos componentes integrados para formar el nuevo programa.

Conclusiones y recomendaciones parciales

Los dos programas revisados en este capítulo, si bien tienen antecedentes, fueron reestructurados en 2011 y al siguiente año se les aplicó una ECR con los mismos TR genéricos diseñados por el Coneval, cuyo desarrollo mostró que hubiera sido más adecuado aplicar sólo una evaluación de diseño y postergar la ECR para cuando el programa tuviera mejor sistematizada su información. Por otra parte, la Sagarpa insiste en que los TR son muy acotados y no le brindan información útil a su proceso decisorio, por lo que proponen abrir esta discusión no sólo con el Coneval y las diversas secretarías del gobierno federal, sino también con evaluadores externos y académicos.

Los equipos evaluadores tuvieron alto nivel de formación, adecuada composición interdisciplinaria y amplia experiencia en evaluación. Uno de ellos fue contratado mediante adjudicación directa con un costo de \$730 000.00 (incluyendo un documento extra) y el otro fue parte de un convenio de colaboración técnica preexistente que permitió abatir sensiblemente los costos a \$347 724.77.

En los dos casos los equipos evaluadores respondieron a las 51 preguntas de los TR en forma fundamentada con información institucional pero, como ya se mencionó, enfrentaron problemas de escasez y sistematización de la misma debido a la limitada antigüedad del programa, lo que se trató de subsanar aplicando entrevistas a informantes clave.

Los informes finales están adecuadamente estructurados, su redacción es clara y su tamaño muy similar y suficiente (174 y 182 páginas). Ambos se encuentran disponibles en la página web del Coneval y dieron lugar al establecimiento de dos y un ASM respectivamente, pero se desconoce si se les ha dado seguimiento. En el caso de Procampo se menciona que también permitió realizar otros ajustes al programa.

Aunque se echan de menos algunos apartados (metodología, problemas enfrentados en el proceso, etcétera), esto se debe fundamentalmente a las restricciones que imponen los TR, por lo que se juzga que los trabajos son de buena calidad.

Las recomendaciones que se proponen a los problemas específicos identificados incluyen: *a*) emitir TR flexibles y adecuarlos a la problemática concreta del programa a evaluar; *b*) emitir siempre convocatorias dirigidas al menos a

tres instituciones; *c*) aplicar evaluaciones de diseño al inicio del programa y postergar un año las ECR; y *d*) informar siempre sobre la respuesta institucional a la evaluación y el cumplimiento de ASM, desglosadas para cada evaluación realizada.

CAPÍTULO XX
Metaevaluación de evaluaciones
del Gobierno del Distrito Federal

Talina Olvera Mejía
Angélica Rosas Huerta

**Evaluaciones en áreas de política social,
transporte y salud 2010-2011**

Debido al conjunto de similitudes que presentan las evaluaciones realizadas por Evalúa DF y para no hacer reiteraciones innecesarias, la siguiente metaevaluación se aplica a las cuatro evaluaciones realizadas por dicho Consejo en 2010 y 2011:

- Evaluación de diseño e instrumentación de la política de Equidad de Género en el Distrito Federal, 2010.
- Evaluación de diseño e instrumentación de la política de Transporte Público Colectivo de Pasajeros en el Distrito Federal, 2010.
- Evaluación de diseño, instrumentación y resultados de la Política de Participación Ciudadana en el Desarrollo Social del Distrito Federal, 2010.
- Evaluación de operación, resultados e impactos de los Servicios de Salud Materno-Infantiles del Gobierno del Distrito Federal 2006-2010, 2011.

Contexto de las evaluaciones

En 1996 el Congreso aprobó reformas constitucionales para permitir a los ciudadanos del Distrito Federal elegir a sus representantes, y con el cambio de

gobierno la política social fue logrando un mayor impulso, especialmente desde la aprobación de la LDSDF en el 2000, que constituye el marco normativo para el diseño, implementación y evaluación de las políticas sociales.¹

Como se anticipó en el capítulo 8, dichas políticas y programas sociales comenzaron a ser evaluadas desde la integración del Evalúa DF en 2008. Las evaluaciones que a continuación serán objeto de metaevaluación se desarrollaron en 2010 y 2011, durante la segunda mitad del periodo de gobierno del Distrito Federal 2006-2012. El Programa de Desarrollo Social de este gobierno (PDSDF) tuvo como objetivo general:

Desarrollar una política social tendencialmente universal, integral, transversal, participativa y territorializada que permita avanzar en el goce efectivo de los derechos económicos, sociales, culturales y ambientales de los habitantes de la ciudad para abatir la desigualdad, promover la equidad social y de género, desarrollar la inclusión y la cohesión social en la perspectiva de la acumulación de derechos y la continuación en la construcción de un régimen y un estado social de derechos en la Ciudad de México (GDF, 2007:21).

Sus objetivos específicos incluyeron la atención a temas como la educación, la salud, el empleo y los derechos laborales, la vivienda, el acceso al agua potable, la desigualdad de género, los jóvenes, la discriminación, el derecho a la ciudad, la participación ciudadana y el acceso a las tecnologías de información y a la internet, entre otros.

Cuando Evalúa DF decidió pasar de la evaluación de programas a la evaluación de políticas, el objetivo fue concentrarse en temas de carácter estratégico.² Las políticas sociales evaluadas en 2010 incluyeron las áreas de: transporte público colectivo de pasajeros, participación ciudadana en el desarrollo social y equidad de género; y las de 2011 evaluaron los servicios de salud materno-infantil. Los cuatro temas sometidos a evaluación forman parte de las principales acciones y

¹ Para profundizar más en el tema de evaluación de la política social en el Distrito Federal, véanse Cardozo (2010 y 2012b) y Osuna *et al.* (2011).

² Entrevista realizada el 14 de enero de 2014 al maestro Pablo Yanes Rizo, actual funcionario de la CEPAL, quien fungió como director del Consejo de Evaluación del Desarrollo Social del Distrito Federal durante el periodo analizado.

líneas programáticas de esa administración, establecidos en la normatividad y el Programa de Desarrollo Social de este gobierno.

El transporte público colectivo de pasajeros (TPCP) se ha caracterizado desde hace varias décadas por brindar un servicio inadecuado, con unidades en mal estado, prepotencia y falta de capacitación de los conductores, desorden en rutas y ramales, y un escaso control por parte de la autoridad, ocasionando inseguridad para el usuario, saturación de vialidades, aumento en el tiempo promedio de los viajes y alto impacto ambiental. Si bien las tarifas subsidiadas cumplen una función social por su bajo costo, no permiten la renovación de unidades ni mejorar las condiciones de vida de los conductores. El servicio de taxis es una de las flotas de transporte más numerosa en que, además de las unidades registradas, hay otro tanto que opera de manera irregular, lo que ha generado inseguridad entre los usuarios. El exceso de vehículos particulares que sólo traslada un porcentaje bajo de pasajeros es otro de los graves problemas que afectan seriamente la movilidad de la ciudad.

Ante esto, la reciente política de transporte público del GDF se ha caracterizado por la puesta en marcha de importantes estrategias con el objetivo de fortalecer y privilegiar el transporte público y la movilidad en la ciudad, al tiempo que desincentiva la movilidad mediante el automóvil particular, taxi o micro buses. Esta política busca también que el traslado de los pasajeros sea eficiente, cómodo, económico y seguro. Los medios de transporte público están conformados por el Metro, el Sistema de Transporte Eléctrico (STE), la Red de Transporte de Pasajeros (RTP), el Metrobús y los corredores de transporte público.³ La evaluación busca contribuir a que dicha estrategia beneficie a los sectores de la población en condiciones de pobreza y vulnerabilidad. En los últimos años han aparecido servicios de movilidad alternativa vinculada a la promoción de un nuevo orden urbano, como los ciclo-taxis en el Centro Histórico de la Ciudad asistidos con tecnología híbrida, las ciclo-vías y ciclo-estacionamientos como infraestructura de transporte sustentable, y el sistema Ecobici que cuenta con ciclo-estaciones en varias colonias de la ciudad.

La participación ciudadana está poco desarrollada (capítulo 9) y constituye uno de los ejes de la política social del GDF, por lo que existen acciones y pro-

³ Los centros de transferencia modal se han creado para agilizar el transbordo de los diferentes medios de transporte de manera segura y rápida, sin interferir en las vialidades.

gramas dirigidos a fomentarla. Éstos buscan institucionalizar los procedimientos que la garanticen para alcanzar en el largo plazo una democracia participativa consolidada y con instrumentos efectivos que permitan otorgarle carácter vinculatorio a sus decisiones.

La institucionalización de la participación ciudadana en el ámbito público se dio con la promulgación de la Ley de Participación Ciudadana (1998), de la LDSDF y la de Fomento de las Actividades de Desarrollo Social de las Organizaciones Civiles (ambas del 2000), con lo que se abrieron espacios de participación a organizaciones civiles y sociales (consejos consultivos en las secretarías, consejos de participación ciudadana en los programas de gobierno, foros y procesos de consulta, participación en redes vecinales, planeación participativa y colaboración en programas específicos). Esta participación se mantuvo en el Programa General de Desarrollo del Distrito Federal 2001-2006 y en 2004 se introdujeron en la Ley de Participación Ciudadana la rendición de cuentas, la red de contralorías ciudadana y la asamblea ciudadana, reconociendo a los comités y los consejos ciudadanos.

Aunado a esto se han creado programas participativos como el Programa de Coinversión para el Desarrollo Social del Distrito Federal en 2001, el Programa Comunitario de Mejoramiento Barrial de 2007 o el de Financiamiento a la Asistencia e Integración Social de 2001. En el Programa de Apoyo a los Pueblos Originarios la asamblea comunitaria debe avalar los proyectos y las actividades en las que se ejercerá el presupuesto. Finalmente, el Programa de Unidades Habitacionales Ollin Callan apoya a la organización condominal de los conjuntos habitacionales.

La desigualdad de género es otro de los problemas relevantes de la ciudad, por lo que desde 1997 se busca promover su equidad, institucionalizando la perspectiva correspondiente en la cultura organizacional de las dependencias y en todas las acciones y niveles de gobierno. En 1998 se preparó el primer Plan de Igualdad de Oportunidades para las Mujeres del Distrito Federal, y en 2002 se creó el Instituto Nacional de las Mujeres del Distrito Federal (Inmujeres-DF)⁴ como organismo público descentralizado.

⁴ Su antecedente desde 1999 fue el Instituto de la Mujer del Distrito Federal como organismo desconcentrado de la Secretaría de Gobierno.

Pero fue la administración 2006-2012 la que desarrolló iniciativas legislativas y administrativas y puso en marcha programas emblemáticos: en 2007, la reforma al Código Penal en materia de Interrupción Legal del Embarazo⁵ (ILE), la modificación correspondiente a la Ley de Salud del Distrito Federal y la promulgación de la Ley de Igualdad Sustantiva entre Mujeres y Hombres en el Distrito Federal (LISMHDF); en 2008, la promulgación de la Ley de Acceso de las Mujeres a una Vida Libre de Violencia en el Distrito Federal y la creación de la Comisión Interinstitucional de Género del Distrito Federal para dar cumplimiento a la LISMHDF y a la planeación del presupuesto con enfoque de género; y en 2009, la instalación del Programa General de Igualdad de Oportunidades y No Discriminación hacia las Mujeres en la Ciudad de México.

Respecto al último caso, el derecho a la salud se encuentra consagrado en el artículo cuarto constitucional y México, como miembro de la OMS, está obligado a emprender acciones para lograr que todos tengan acceso a los servicios sanitarios. En ese contexto, la Ley de Salud del Distrito Federal señala que el GDF tiene la atribución de prestar los servicios integrales de atención materno-infantil que comprenden, entre otros, la atención de la mujer durante el embarazo, parto y puerperio. La Secretaría de Salud del Distrito Federal (SSDF) es la responsable directa de la operación de los hospitales de materno-infantil.⁶

Si bien las principales causas de muerte en el Distrito Federal se encuentran relacionadas con las enfermedades del corazón, la diabetes y el cáncer, también existen datos preocupantes sobre la mortalidad materno-infantil, en un contexto nacional que indica que México, como país, no será capaz de cumplir el Objetivo del Milenio en la materia. Al respecto, en 2008, la SS y el Instituto Nacional de Estadística Geografía e Informática (INEGI) reportaron que el Distrito Federal tenía el cuarto lugar nacional en el número de defunciones maternas, y el primer lugar en muertes fetales. Los datos sorprendieron dado el nivel relativamente alto del desarrollo de la infraestructura hospitalaria y los recursos médicos en

⁵ En 2002 ya se habían ampliado las causales del aborto permitido para incluir las alteraciones genéticas o congénitas del producto, la violación y la tipificación del delito de violación por parte del cónyuge, y se había establecido al Ministerio Público como el organismo facultado para autorizarlo en caso de violación.

⁶ En total siete al momento de hacerse la evaluación.

el Distrito Federal, por lo que fue pertinente cuestionarse sobre las causas que explicaban esta situación.

Frente a este panorama, y con el antecedente de haber realizado en 2008 una evaluación del diseño y operación del Programa de Acceso Gratuito a los Servicios Médicos y Medicamentos, Evalúa DF decidió realizar una evaluación de operación, resultados e impacto de los servicios de salud materno-infantil que el GDF brinda a las mujeres en edad reproductiva, particularmente durante el embarazo, parto y posparto, así como a los niños, cuyos problemas al nacer extienden sus secuelas principalmente durante el primer año de vida. Fue el primer y único esfuerzo del Evalúa DF por realizar una evaluación de este tipo.

Marco normativo

El marco normativo que sustenta las evaluaciones de la política social en el Distrito Federal lo conforma la LDSDF y su Reglamento, así como el Estatuto Orgánico del Evalúa DF. La LDSDF señala la importancia de la evaluación en su artículo 42:

[...] las evaluaciones constituyen procesos de aplicación de un método sistemático que permite conocer, explicar y valorar al menos, el diseño, la operación, los resultados y el impacto de la política y programas de Desarrollo Social. Las evaluaciones deberán detectar sus aciertos y fortalezas, identificar sus problemas y en su caso, formular observaciones y recomendaciones para su reorientación y fortalecimiento.

Características que fueron retomadas por los términos de referencia de las evaluaciones analizadas.

La normatividad mencionada establece las facultades que tiene Evalúa DF en la materia y describe con claridad todo el proceso de realización de las evaluaciones externas, desde la convocatoria hasta la emisión de recomendaciones y las posibles controversias que éstas pudieran generar.⁷ Como dato importante, las recomendaciones del Evalúa DF tienen carácter vinculatorio.

⁷ Los procedimientos se encuentran descritos detalladamente en Cardozo (2012)

Como se mencionó en el capítulo 9, los TR que acompañan a cada convocatoria emitida por este Consejo no son genéricos sino que especifican de manera particular los requerimientos para cada una de sus evaluaciones.⁸ Todos señalan que serán externas y sólo la de servicios de salud materno-infantil menciona expresamente que se trata de una evaluación de carácter intermedio; sin embargo, esto se sobreentiende en la medida en que no son iniciales ni mencionan la posibilidad de terminar el programa. Ninguna de las evaluaciones se enfoca en programas, tres de ellas lo hacen en políticas y una más en servicios.

Sus objetivos pretenden contribuir a la rendición de cuentas; pero principalmente a la mejora de la propia política o servicio, lo que sólo se omite en la evaluación de la política de participación ciudadana. En cada uno de los casos se precisa el alcance de la evaluación (contenido o fase de la política que se evalúa), sin embargo, algunos incluyen ámbitos que rebasan la etapa considerada.⁹ Todos son muy ambiciosos: tres plantean la evaluación del diseño y la instrumentación, una agrega los resultados, y la cuarta solicita la operación, resultados e impacto,¹⁰ pero no se trata sólo de evaluar las acciones sino de ubicarlas en el marco de sus antecedentes. Tanto sus objetivos específicos como sus definiciones mantienen una alta coherencia y relación con sus objetivos generales. Incluyen una sección en que indica detalladamente los temas solicitados, pero no se explicitan los criterios de valor para hacerlo, los que se consideran esenciales. Sorprende que, dada la importancia y arraigo que se espera vaya teniendo la evaluación, ninguno de los TR retoma de la normatividad general del Evalúa DF el propósito de

⁸ Entrevista al maestro Yanes Rizo, ya mencionada.

⁹ Por ejemplo: *a*) en el caso de la política de TPCP al mencionar que “[...] la evaluación de la política del TPCP en la Ciudad de México se orienta a analizar su diseño e instrumentación con énfasis en los efectos sociales que producen entre sus habitantes” (p. 3); *b*) en la política de equidad de género que en su tercer objetivo específico menciona: “Analizar y evaluar las características y el grado de concreción efectivo que se ha logrado en la transversalización de la perspectiva de género y en la naturaleza integral de la política” (p. 4), siendo que se limita a la evaluación del diseño y la instrumentación y que además solicita un diagnóstico propio, que tampoco es el foco de la evaluación; y *c*) en la política de participación ciudadana al incluir en el análisis de la implementación “[...] el logro de sus metas y la valoración de los resultados alcanzados en función de los recursos utilizados y las metas previstas (eficacia y eficiencia)” (p. 4).

¹⁰ Esta última evaluación contó con recursos mucho más significativos que las demás.

desarrollar dicha función dentro de los organismos involucrados o dentro de las propias políticas y programas evaluados.

Concretamente, dado que la evaluación de la política de TPCP y equidad de género se enfocan en el diseño e instrumentación, ambas presentan requerimientos muy similares: antecedentes, pertinencia y suficiencia del diagnóstico, grado de armonización normativa, análisis de documentos rectores, análisis del presupuesto, evaluación exhaustiva de algún programa clave o relevante. En el caso de participación ciudadana, que además solicita la evaluación de resultados, se agrega la eficacia y eficiencia de los mecanismos de participación. Los servicios de salud materno-infantil cuya evaluación cubre la operación, los resultados y el impacto, solicitan analizar los recursos disponibles, valorar los servicios, la aplicación de la perspectiva de género, el avance en el cumplimiento del derecho a la protección de la salud, el grado de satisfacción de los usuarios y las causas de la alta incidencia de la mortalidad materna anual. En general, estos requerimientos son pertinentes con la fase de la política que se espera evaluar, sin embargo, como mencionamos, en algunos casos se incluyen temas que la rebasan.

A pesar de la dificultad para delimitar políticas de manera clara, se incluyen definiciones de las mismas que contribuyen a identificar los elementos que la conforman (normas, acciones, medidas, programas y omisiones) y que se espera sean evaluados. A partir de la lectura completa de los TR se ubican de manera más específica los textos legales, documentos rectores, programas u hospitales, cuya lista es amplia.

Todos los TR especifican de manera clara y completa los contenidos que deben incluir las propuestas técnicas. En ellos sobresale el marco teórico y metodológico, además de las fuentes de información, un cronograma de trabajo, los apartados básicos en la estructura del informe y un resumen curricular del equipo evaluador. En el caso de la política de participación ciudadana y los servicios de salud materno-infantil se especificó un poco más sobre el contenido que debería incluir el marco teórico y metodológico. Mención especial merecen los TR de los servicios de salud materno-infantil que dentro del apartado de la propuesta técnica también incluyeron la metodología y técnicas de investigación, haciéndolas mucho más claras y específicas.

También todos solicitaron tres informes. Los aspectos que deben cubrirse en el informe final son bastante completos y pertinentes, incluyendo conclusiones

y recomendaciones, así como el marco teórico y metodológico que lo sustenta. En el caso de la política de TPCP no se señala un apartado de conclusiones, sólo la valoración global de la política, que se solicita en todos. Los TR de la política de participación ciudadana precisan que ésta se debe enmarcar en la discusión teórica realizada. En cuanto a las recomendaciones, todos las mencionan como un apartado a incluir pero sólo los relativos a la política de participación ciudadana las solicitan para cada uno de los programas analizados, los componentes y la política en general. Con esto se hace explícito que las evaluaciones deben contener los elementos que le dan ese carácter al trabajo realizado: valoración y recomendaciones.

Equipo evaluador y recursos económicos

La convocatoria fue clara respecto a quienes podían participar en los trabajos, manteniendo un alto grado de profesionalismo entre los evaluadores al convocar a profesores y/o investigadores adscritos a instituciones de educación superior, de investigación científica, profesionales pertenecientes a organizaciones civiles o sociales sin fines de lucro, y/o independientes, expertos en política social y preferentemente en los temas relacionados con la evaluación. Para los servicios de salud materno-infantil se especificó que también contarán con experiencia en evaluación de impacto. Todos debían estar inscritos en el directorio de evaluadores externos del Evalúa DF, tal y como lo señala la normatividad. La composición efectiva lograda fue la siguiente:

El equipo evaluador de la política de TPCP, adscrito al Programa Universitario de Estudios sobre la Calidad de la UNAM, incorporó especialistas en geografía, ingeniería, derecho, economía, sistemas, matemáticas aplicadas, y ciencias atmosféricas. El coordinador fue el único que un año atrás había tenido una experiencia de evaluación en Evalúa DF, en el caso de la política de acceso al agua potable. Tiene una formación en antropología social, una maestría en estudios regionales y un doctorado en sociología en la Escuela de Altos Estudios en Ciencias Sociales de París. Sus líneas de investigación abarcan la geopolítica del agua, los recursos hídricos, la desigualdad y segregación urbana y la planeación participativa.

La evaluación de la política de participación ciudadana estuvo coordinada por una licenciada en trabajo social, con maestría en investigación y ciencias sociales y doctorado en diseño urbano por la UAM-Azcapotzalco, así como experiencia en temas de participación ciudadana y democracia. Una de las investigadoras de su equipo tiene licenciatura y un doctorado en sociología por la UNAM. Otra cuenta con una maestría, su área de interés es el derecho a la ciudad y es miembro activo en asociaciones profesionales que se dedican a la investigación sobre movimientos urbanos. Debido a lo anterior, cuenta con experiencia en investigación sobre programas cofinanciados y que se enfocan en el trabajo conjunto con organizaciones sociales, incluso ha colaborado en trabajos previos para Evalúa DF con otros equipos de evaluación.

La responsable de la evaluación de la política de equidad de género también tiene una licenciatura en trabajo social, además de un magister en gestión y políticas públicas por la Universidad de Chile; cuenta con experiencia en el tema de género, y ya había realizado trabajos de consultoría y evaluación de programas públicos. Su equipo tenía formación y experiencia en el área de las políticas públicas, derechos humanos, gestión y evaluación de proyectos.

Los TR para la evaluación de los servicios de salud materno-infantil fueron los únicos que hicieron explícita la solicitud de equipos interdisciplinarios. La evaluación estuvo dirigida por una consultora con formación en sociología y una maestría en economía por El Colegio de México, que había fungido como directora de Evaluación en una dependencia de la Sedesol. El trabajo de campo fue coordinado por una especialista con formación académica en psicología y experiencia como consultora en temas de opinión pública, enfocada en prestar sus servicios al gobierno federal y los estatales.

La interdisciplinariedad del equipo evaluador es una cualidad necesaria que se pudo encontrar entre quienes realizaron los trabajos, puesto que el Consejo tenía claro que entre los equipos debía existir un equilibrio entre el conocimiento de la metodología de la evaluación y del tema a evaluar. Todas las evaluaciones tuvieron un coordinador con posgrado y experiencia en el tema; en algunos casos también en evaluación. Los perfiles son adecuados a las necesidades de evaluación.

En cuanto al financiamiento de cada una de las evaluaciones, en el caso del Evalúa DF éste proviene de recursos presupuestales independientes de las instancias involucradas en la política o servicios evaluados. Los costos de cada

trabajo se especificaron e hicieron públicos en sus respectivas convocatorias: \$348 000.00 para las realizadas en 2010 y \$1 050 000.00 en 2011. La diferencia se justifica debido a que la última incluyó la evaluación de impacto con trabajo de campo. Ambos montos se consideran adecuados al trabajo solicitado.

Procesos de trabajo

Las evaluaciones siempre están sujetas al ciclo presupuestal anual y se definen en el Programa Anual de Evaluaciones que se acuerda entre los consejeros y se publica a más tardar en marzo. Una vez recibido el presupuesto definitivo, se elaboran y aprueban los TR y se emiten las convocatorias. En el caso de las evaluaciones de 2010 y 2011 el Consejo de Evaluación emitió las convocatorias para las evaluaciones en abril y marzo, respectivamente. Éstas fueron enviadas a todos los integrantes del directorio del Evalúa DF.

La convocatoria de 2011 fue la única que estableció formalmente una práctica que se había ido generando en el sentido de realizar sesiones aclaratorias de las convocatorias¹¹ y consideró la realización de reuniones previas con autoridades del sector para intercambiar puntos de vista sobre la evaluación propuesta, facilitar la entrega de información y acordar las entrevistas a realizar.

Salvo la evaluación para la política de equidad de género que se consideró desierta y fue necesaria la publicación de una segunda convocatoria que recorrió un mes los plazos originalmente establecidos, todas las demás se cubrieron en los tiempos previstos.

La asignación definitiva de los contratos entre quienes respondieron en tiempo y forma a la convocatoria fue decidida por el pleno del Consejo tras analizar la propuesta técnica, su apego a los TR, la calidad del equipo evaluador (conocimiento y experiencia), su solvencia ética, la ausencia de posibles conflictos de interés, la calidad y el contenido de los productos a entregar. Entre la fecha límite para presentar la propuesta y la publicación de los resultados transcurrieron como máximo dos semanas. En algunos casos se realizaron sugerencias a las propuestas de los equipos evaluadores.

¹¹ Entrevista al maestro Yánes Rizo, ya mencionada.

Todos los equipos evaluadores contaron con el acompañamiento de un consejero durante la realización de la evaluación. Su propósito fue asegurar que la información requerida por los evaluadores les fuera proporcionada oportunamente, y en su caso, solucionar dudas que surgieran durante el trabajo, para garantizar la entrega de los informes en tiempo y forma (Cardozo, 2012b).

Los informes finales de evaluación de las políticas de TPCP y de servicios de salud materno-infantil señalan que la duración de la evaluación fue de cuatro y seis meses respectivamente, pero en la práctica todas las evaluaciones excedieron el plazo fijado,¹² en parte debido a las correcciones sugeridas, para las que originalmente se había destinado un mes de trabajo (diciembre). Este mes adicional no se dispuso en el caso de la evaluación de equidad de género debido a su retraso inicial (Cuadro 27). El informe de la política de TPCP señaló que el tiempo de trabajo fue insuficiente para alcanzar resultados categóricos.

Metodología de la evaluación

A excepción del informe sobre la evaluación de la política de participación ciudadana, los demás incorporan las técnicas de recolección de datos y hacen explícitas las fuentes de información. Los informes de TPCP y de equidad de género incluyen también el procedimiento que han recorrido para elaborar las recomendaciones, incluso TPCP agrega el enfoque interdisciplinario que guía el trabajo. La evaluación más completa (TPCP) menciona el tipo de trabajo, el área responsable y la acción a realizar para cubrir cada uno de los requerimientos de los TR. El proceso constó de cuatro fases, la primera fue de caracterización y diagnóstico, la segunda de la evaluación sustantiva y pronóstico, la tercera de la valoración global, y la última correspondió a las recomendaciones y observaciones. Estas fases contienen las dos actividades básicas de toda evaluación, mismas que se reflejan en el trabajo. Al tiempo que se van detallando cada una se nota una adecuación total del diseño metodológico respecto al contexto y al objeto a evaluar.

¹² En las evaluaciones de TPCP y participación ciudadana se continuaba realizando entrevistas en febrero de 2011 y el informe final de la última estuvo listo recién en mayo de ese año.

CUADRO 27
Distribución del tiempo de trabajo por evaluación y etapas

	Transporte público	Participación ciudadana	Equidad de género	Salud materno-infantil
Emisión de la convocatoria	Abril	Abril	Abril	Marzo
Plazo máximo de presentación de propuestas	7 de mayo de 2010	7 de mayo de 2010	18 de junio de 2010	8 de abril de 2011
Publicación de resultados	C: 1 de junio de 2010 E: 18 de mayo de 2010	C: 1 de junio de 2010 E: 18 de mayo de 2010	C: 30 de junio de 2010 E: 30 de junio de 2010	C: 15 de abril de 2011 E: 12 de abril de 2011
Inicio de trabajos	Agosto de 2011	Agosto de 2011	ND	Junio de 2011
Informe inicial	C: 30 de junio de 2010 Cumplido	C: 30 de junio de 2010 Cumplido	C: 30 de julio de 2010 ND	Un mes después de firmado el contrato Cumplido
Informe intermedio	C: 30 de septiembre de 2010 Cumplido	C: 30 de septiembre de 2010 Cumplido	C: 31 de octubre de 2010 ND	90 días después del primer informe Cumplido
Borrador final	C: 30 de noviembre de 2010 Cumplido	C: 30 de noviembre de 2010 ND	C: 30 de diciembre de 2010 ND	C: 30 de noviembre de 2011 ND
Informe final	C: 31 de diciembre de 2010 E: 25 de febrero de 2011	C: 31 de diciembre de 2010 E: Mayo de 2011	C: 31 de diciembre de 2010 E: 28 de febrero de 2011	C: 31 de diciembre de 2011 E: 2011

C: Fecha límite establecida en la convocatoria; E: Fecha de entrega efectiva.

Fuente: elaboración propia a partir de las convocatorias, términos de referencia e informes de evaluación de los cuatro casos.

En la evaluación de equidad de género sólo se mencionan las etapas sin mayor detalle: revisión de documentación oficial e histórica, trabajo de campo, interpretación de resultados, redacción de informes, e incorporación de contribuciones para el análisis final. En este proceso están ausentes dos etapas necesarias en toda evaluación: la valoración de la información y la explicitación de recomendaciones. La metodología usada fue básicamente cualitativa, la cual es adecuada a las necesidades de esta evaluación.

En el caso de la política de participación ciudadana no se describe el procedimiento de evaluación, sólo se explicitan las preguntas, dimensiones y variables a considerar, lo que da claridad al proceso. Por último, en el caso de los servicios de salud materno-infantil, aunque el informe incluye un capítulo metodológico, éste se limita básicamente a los instrumentos de recopilación de información, por lo que no queda claro si se aplicó efectivamente la metodología general propuesta por los términos de referencia.

Las técnicas más utilizadas para la recolección y el análisis de información fueron el análisis documental y las entrevistas (Cuadro 28). Todas las evaluaciones realizaron análisis documental, pero TPCP y salud materno-infantil usaron de manera relevante datos provenientes de evaluaciones y encuestas de distintas fuentes. Salvo este último caso, las demás realizaron también entrevistas, pero es la evaluación de participación ciudadana la que más hace referencia a los hallazgos surgidos de éstas, TPCP las refiere con poca frecuencia, pero ambas citan la fuente y en ocasiones también la fecha de realización. En cambio la política de equidad de género las integra en pocas ocasiones y sin agregar datos sobre el entrevistado.

Para la evaluación de los servicios de salud materno-infantil se realizaron encuestas, observación *in situ*, y grupos focales, esto debido a los requerimientos de evaluación y de los TR. En el informe se suele hacer constante referencia a la información obtenida de ellas.

Aunque la política de equidad de género menciona que realizó observación directa, la información obtenida mediante ésta no se refiere dentro de la evaluación, y aunque en un inicio consideró la aplicación de cuestionarios esto no fue posible debido a la recepción tardía de los listados de funcionarios involucrados.

Las evaluaciones de la política de TPCP y de servicios de salud materno infantil exponen la justificación de la elección de sus instrumentos en relación

con el propósito del trabajo. Especialmente la segunda menciona los criterios para la determinación del tamaño muestral y su selección y hace una descripción detallada de cada una de las técnicas. Ambas hacen explícita la información que analizan mediante trabajo de gabinete y de campo y su diseño se muestra suficiente para la obtención de buenos resultados y útil para la evaluación, además de su rigurosa aplicación posterior de acuerdo con lo planeado.

La evaluación de la política de equidad de género no explica su selección pero es la única que adjunta los guiones específicos para cada una de las entrevistas. En cambio, en todos los casos se incluye el listado de personas entrevistadas, con su nombre y cargo, y en el de la política de transporte público se señala también la fecha de la entrevista. Generalmente no en el apartado metodológico, pero sí en anexos (TPCP) o al momento de su análisis (equidad de género) se incluye también el listado de las demás fuentes consultadas.

En todos los casos, las fuentes y técnicas de evaluación se ajustan a los TR y son adecuadas aunque no siempre suficientes respecto al contexto y objeto a evaluar. Esto último debido a que en general las evaluaciones con contenido de implementación sólo toman en cuenta la visión de los funcionarios responsables o gestores, descuidando la percepción de los beneficiarios porque considerarlas hubiera implicado hacer un trabajo de campo no previsto en función de los recursos disponibles. La evaluación de los servicios de salud materno-infantil es la única cuyos TR la solicitaron y que incorporó ambas visiones, mientras la relativa a TPCP supo subsanar sus limitaciones para lograr un diseño metodológico e instrumentos que se adecuaron a las necesidades de evaluación sin demeritar sus resultados enmarcándolas en el derecho a la movilidad. Por su parte, la evaluación de la política de participación ciudadana, si bien no incluyó a los beneficiarios, involucró a un número importante de personas de la sociedad civil y organizaciones no gubernamentales (ONG) a partir de las entrevistas. Finalmente, en el caso de la política de equidad de género, a pesar de que el informe menciona que la metodología cualitativa permite analizar aspectos desde la perspectiva de los actores y que contempló entrevistas con algunas representantes de la sociedad civil, sólo logró efectuar dos a beneficiarias de los servicios.

CUADRO 28
Instrumentos de recopilación de información utilizados

Política	Recopilación documental	Entrevistas	Encuestas	Observación <i>in situ</i>	Grupos focales	Bases de datos de otras fuentes
Participación ciudadana*	x	x				
Transporte público	x	x				x
Equidad de género	x	x				
Salud materno-infantil	x		x	x	x	x

* El apartado relativo a metodología no menciona los instrumentos de información, pero éstos se identifican durante la lectura del informe.

Fuente: elaboración propia a partir de los informes de las evaluaciones consideradas y de la propuesta metodológica para la evaluación de la política de participación ciudadana.

Informes finales

Los informes finales de evaluación se adecuaron al número máximo de páginas solicitado en los TR (400): el de la política de participación ciudadana es el más extenso con 395 páginas, seguido por el de TPCP, el de equidad de género y el de los servicios de salud materno-infantil con 284, 233 y 186 páginas respectivamente. Por la amplitud y complejidad de su tema se hubiera esperado que el último fuera más extenso.

Los índices y el contenido de los cuatro informes muestran la información esencial con una estructura y secuencia lógica en su presentación. El grado de claridad y suficiencia en cada uno de los puntos varía en cada evaluación. Cabe destacar que el informe de evaluación de la política de equidad de género es el único que incluye el resumen ejecutivo solicitado, éste contiene básicamente la misma información de la introducción, agregando las conclusiones y recomendaciones de manera íntegra.

Tanto el informe de evaluación de la política de TPCP como el de equidad de género, que abarcan el diseño y la instrumentación, cubren los aspectos de

antecedentes, pertinencia y suficiencia del diagnóstico, grado de armonización normativa, análisis de documentos rectores y del presupuesto, y evaluación de algún programa clave. Dicho contenido resulta suficiente y adecuado a los requerimientos de la evaluación. La evaluación de la política de TPCP adecuó las solicitudes de los TR, tomándolas como guía y respondió con un alto contenido valorativo. Todos los apartados guardan relación y coherencia entre ellos, por lo que reflejan de manera clara el análisis realizado previamente. Incluye un capítulo de prospectiva y desafíos estratégicos con propuestas para guiar acciones futuras.

En la evaluación de la política de equidad de género la información, presentada a manera de hallazgos,¹³ responde a las necesidades de evaluación tal cual se presentaron en los TR. La evaluación se limita prácticamente a describir las acciones relevantes realizadas en cada una de las dimensiones solicitadas sin acompañarse, en la mayoría de los casos, de un análisis o valoración. Ante la diversidad de temas a valorar, en muchos de los casos la información suele ser superficial, aunque en algunos se incorpora una descripción más detallada, seguida de una valoración y alguna sugerencia. Los juicios, generalmente se acompañan de la frase “a juicio del equipo evaluador”, con lo que pareciera que las valoraciones, y posteriores recomendaciones, están sustentadas sobre todo en las percepciones de los evaluadores y criterios no explícitos.

Como se recordará, la evaluación de la política de participación ciudadana cubre los resultados, además de los aspectos de diseño e implementación. El informe analiza cuatro componentes centrales: los instrumentos institucionalizados para la participación ciudadana, el funcionamiento de los programas orientados a la promoción y garantía de la participación ciudadana en el desarrollo social, las agendas y demandas de participación de los actores y los desafíos de la participación ciudadana en el ámbito territorial. Con base en éstos se valoran congruencia, consistencia interna, pertinencia, suficiencia, desempeño, claridad de funciones, manejo de recursos, y contrastación de los propósitos con lo logrado, por mencionar algunos. Esto permite mantener una alta coherencia y adecuación entre los aspectos evaluados y solicitados en los TR, además de ser suficientes en razón del contenido de la evaluación. Los hallazgos se presentan en función de los componentes básicos, dedicando un capítulo a cada uno de ellos. El componente valorativo también suele estar presente a lo largo de la evaluación.

¹³ Llegan a contabilizarse 77 hallazgos.

A su vez, la evaluación de los servicios de salud materno-infantil cubre también el impacto. Al igual que en dos de los casos anteriores, sin descuidar los TR, se adecuan las peticiones para lograr una presentación más clara. Los aspectos contemplados en la evaluación cumplen con los temas solicitados en los TR. En este sentido, respecto a la operación se analizaron los recursos presupuestales, los humanos, la infraestructura, los procesos de atención (acceso, tiempo y condiciones de espera, tiempo de atención, información recibida en la consulta, suficiencia y oportunidad de medicamentos, valoración de los procesos de atención). En cuanto a los resultados de los servicios se valoró el acceso universal, la calidad y calidez de los servicios, los factores que explican sus deficiencias, y el derecho de las mujeres a decidir libremente sobre su maternidad. Para el impacto se plantearon los indicadores básicos, los de utilización y cobertura, y los de salud reproductiva. Además se incluyeron también antecedentes, marco legal y estructura orgánica, todos ellos solicitados en los TR. En varios de los puntos el análisis es escaso en profundidad por lo que los resultados no siempre responden plenamente al esfuerzo solicitado para acercarse a la medición de impacto y encontrar explicaciones. De todas formas, estas dos últimas evaluaciones son las más completas en sus contenidos, con información útil y pertinente para la toma de decisiones.

El análisis más detallado muestra que todos los informes comienzan con una introducción, y antes de exponer los resultados de la evaluación incluyen capítulos sobre el marco teórico y el metodológico, para terminar con las conclusiones y recomendaciones. Incluyen también la bibliografía y, en algunos casos, anexos.

En general, las cuatro introducciones retoman los puntos esenciales de los TR (objetivos, metodología, capitulado) e incorporan elementos suficientes para una comprensión de la evaluación, las de TPCP y de participación ciudadana son las más completas. Esta última y los servicios de salud materno-infantil mencionan la dificultad para delimitar sus objetos de evaluación.

Todos los informes incluyen un capítulo teórico-metodológico,¹⁴ cuyo grado de claridad y suficiencia es variable. Los cuatro definen los principales

¹⁴ La referencia metodológica no se incluye en el informe de evaluación de la política de participación ciudadana sino en un anexo, aunque sin especificar las fuentes y técnicas de recolección de información.

conceptos relacionados con el tema a evaluar; en particular discuten sobre las políticas públicas en los casos de TPCP y equidad de género. Aunque, en general, los apartados teóricos son adecuados a las necesidades del tema a evaluar, las fuentes revisadas son más variadas en TPCP y participación ciudadana, lo que les permite exponer de manera más clara, coherente y suficiente las bases teóricas que permean el trabajo realizado y sustentar las posteriores valoraciones. Tanto equidad de género como los servicios de salud materno-infantil sólo retoman una o dos fuentes en cada tema y se limitan principalmente a exponer definiciones, por lo que no es clara la relación entre el enfoque elegido y el trabajo realizado.

La exposición de resultados muestra un nivel bastante alto de congruencia con los TR. Los informes de evaluación de las políticas de TPCP y de participación ciudadana, además de exponer los datos recopilados y las acciones relevantes realizadas, incluyen párrafos dedicados a su valoración. Por el contrario, las otras dos prácticamente se limitan a presentar datos, en algunos casos con tratamientos demasiado breves. En el informe de evaluación de los servicios de salud materno-infantil, el propio equipo evaluador señala que éste no contiene toda la información obtenida y que sólo incluye el trabajo de gabinete y la información generada en el trabajo de campo para los temas requeridos. Agrega que el resto de la información se incorpora en un anexo, sin embargo, el informe no lo contiene. A pesar de que se realizaron encuestas y grupos focales, y de que varios de los apartados se cubren con esa información, ésta se presenta sin mayor análisis ni valoración, cumpliendo de manera muy superficial la mayoría de los requerimientos de los TR, lo cual resta valor al trabajo de campo y a los resultados presentados.

Las cuatro evaluaciones cuentan con apartados exclusivos para conclusiones y recomendaciones, obtenidas mediante la aplicación del marco teórico citado al inicio del trabajo, tal como lo solicitan los TR. En términos generales existe coherencia entre hallazgos, conclusiones y recomendaciones, aunque el grado varía dependiendo de la evaluación, como en los temas previos. Las conclusiones y recomendaciones para las políticas de TPCP y de participación ciudadana son muy similares en su forma de presentación. Ambas señalan que los resultados no son categóricos, sino más bien aproximaciones cualitativas, orientaciones o tendencias, que sientan las bases para profundizar y evaluar posteriormente en mayor detalle los temas abordados. Las conclusiones se apegan a lo solicitado en los TR, retoman el orden expuesto durante el trabajo y resaltan fortalezas,

debilidades y vacíos en los temas demandados en la evaluación, además de incorporar un alto contenido valorativo.

Las recomendaciones cubren todos los aspectos evaluados en consistencia con el análisis realizado y las conclusiones alcanzadas; son puntuales y concretas, identificando a la instancia responsable de ponerlas en práctica. En el caso de la política de TPCP se incluyeron 36 recomendaciones, diferenciando las referidas a diseño y a instrumentación y estableciendo el plazo para cumplirlas; en el de la política de participación ciudadana se emitió por lo menos una recomendación para cada uno de los 50 problemas identificados, agregando la formulación del problema y su causa.

A diferencia de los dos casos anteriores, y a pesar de recuperar en las conclusiones los aspectos más sobresalientes de cada tema y de incorporar las fortalezas y debilidades de la política, en la evaluación de la política de equidad de género se detectan carencias de análisis y valoración. Además, la presentación de las conclusiones y recomendaciones no siguió el orden en que se expuso la información ni contó con subtítulos, por lo que la correspondencia entre temas, conclusión y recomendaciones resultó confusa. La mayoría de las 18 recomendaciones son de carácter declarativo, sin señalar responsable ni plazos, y no retoman todas las que se fueron sugiriendo a lo largo del trabajo.

Las conclusiones de los servicios de salud materno-infantil hacen referencia a los temas abordados, manteniendo el orden de la exposición, sin embargo son en su mayoría descriptivas, con escasas valoraciones. La exposición de sus 30 recomendaciones se hizo por temas y, sin señalar a los responsables de cumplirlas ni sus plazos, resultan comprensibles y reflejan congruencia con el contenido de la evaluación. Pero, las conclusiones y recomendaciones no son suficientes para los requerimientos de una evaluación de impacto, al dejar sin respuesta algunas preguntas fundamentales y limitarse a ofrecer datos sobre la situación actual, dando más la idea de un diagnóstico que de una evaluación de impacto en el periodo 2006-2010.

Los cuatro informes incluyen bibliografía; en tres de los casos se considera pertinente aunque no suficiente, ya que sólo se refiere al tema de la política o servicio considerado. En cambio, en el cuarto (TPCP) se incluye también una amplia y pertinente bibliografía sobre evaluación y políticas públicas. Los informes relativos a las políticas de TPCP y equidad de género incluyen también

anexos, i.e. sobre el análisis de la normatividad y las modificaciones que se consideran necesarias.

Valoración de las evaluaciones

Los cuatro informes de evaluación se caracterizan por un alto grado de homogeneidad en la estructura de su contenido y por la sistematización del trabajo realizado. Esto se debe en gran medida a que los TR así lo solicitan y en todos los casos el apego a ellos ha sido muy elevado. A diferencia de otros trabajos de evaluación, Evalúa DF ha hecho hincapié en que sus evaluaciones se lleven a cabo de manera sistemática, haciendo uso de los métodos de investigación social, emitiendo juicios de valor sobre el desempeño analizado y recomendaciones precisas y bien argumentadas.

Aunque en distintos grados, como ya ha sido señalado, la emisión de los juicios de valor ha estado enmarcada en una adecuada discusión teórica y las evaluaciones han expuesto la metodología que efectivamente han seguido durante la evaluación. Las técnicas de investigación utilizadas han sido adecuadas aunque no suficientes en todos los casos para recoger la percepción de los beneficiarios; sin embargo, así lo solicitaron los TR por carecer de recursos para financiar el trabajo de campo. La sistematización ha permitido que las evaluaciones cuenten con un alto grado de coherencia entre los hallazgos, las conclusiones-valoraciones y las recomendaciones.

En general, el contenido de las tres evaluaciones de 2010 es suficiente y adecuado a las peticiones realizadas en los TR, en el caso de los servicios de salud materno-infantil, al no responder ni presentar de manera clara los resultados esperados en toda evaluación de impacto, se muestra insuficiente. Resalta la buena calidad de las evaluaciones de las políticas de TPCP y de participación ciudadana, que incorporan un alto grado de valoración en los temas analizados y presentan recomendaciones más precisas. Las evaluaciones de la política de equidad de género y de los servicios de salud materno-infantil ofrecen una calidad aceptable al reflejar una sistematización en el trabajo. En el caso de equidad de género el trabajo se apega por completo a los TR pero presenta análisis y valoraciones insuficientes en varios de los temas. En el caso de salud materno-infantil, aunque en lo general intentó dar respuesta a lo solicitado en

los TR y mostró información obtenida de encuestas propias y de grupos de discusión, no cumplió plenamente con los requerimientos y expectativas de una evaluación de impacto.

Finalmente, cabe destacar que el trabajo del Evalúa DF y los TR mantienen un alto grado de coherencia con lo establecido en la LDSDF y el estatuto orgánico del Evalúa DF. Las evaluaciones se han adjudicado a los equipos evaluadores que han cumplido con los requisitos establecidos en las convocatorias, de acuerdo con lo señalado en la normatividad. También ha habido consistencia entre el alcance de las evaluaciones, sus tiempos y presupuestos.

Respuesta institucional a la evaluación y recomendaciones

Los documentos emitidos por las dependencias en respuesta a los resultados de las evaluaciones de transporte,¹⁵ participación ciudadana¹⁶ y equidad de género¹⁷ tuvieron extensiones variables: entre una y 30 páginas. Hubo varios comentarios generales favorables a los esfuerzos de evaluación de políticas y los específicos fueron pertinentes, concretos y claros, aunque en algunos casos rebasaron las posibilidades de la evaluación, como por ejemplo, la sugerencia de considerar más programas y/o documentos para evaluar, o de realizar encuestas piloto dirigidas a los usuarios para que la evaluación también refleje sus puntos de vista. También hubo pronunciamientos en el sentido de agregar acciones no mencionadas en el informe que estarían desarrollando las dependencias, con objeto de superar algunos déficits señalados en la evaluación, y precisar datos

¹⁵ Direcciones Generales del Sistema de Transporte Colectivo, Red de Transporte de Pasajeros del DF, Metrobús; secretarías de Desarrollo Urbano y Vivienda, Finanzas, Medio Ambiente, Inmujeres-DF y Programa Viajemos Seguras en el Transporte Público de la Ciudad de México.

¹⁶ Secretaría de Desarrollo Social por medio de la Dirección General de Igualdad y Diversidad Social y la Dirección General del Instituto de Asistencia e Integración Social.

¹⁷ Inmujeres-DF, Secretaría de Salud, Secretaría de Finanzas (mediante la Dirección de Análisis Programático de Equidad de Género de la Subsecretaría de Egresos) y Procuraduría General de Justicia del Distrito Federal (a partir de la Subprocuraduría Jurídica y de Derechos Humanos).

y actualizar información, no compartir algunas valoraciones u opiniones por considerarlas infundadas, solicitar una modificación en la redacción o agregar un glosario de términos usados.

Ante la atribución del Evalúa DF de emitir recomendaciones de carácter vinculatorio sobre cada política evaluada, y tomando como base las realizadas por el equipo evaluador, en 2010 se emitieron 57 recomendaciones. En relación con la evaluación de los servicios de salud materno-infantil, el Comité de Evaluación y Recomendaciones que solicitó el trabajo a nombre del Evalúa DF realizó su última sesión ordinaria el 22 de mayo de 2011 y en ésta apenas aprobó el envío del informe de evaluación a la SSDF, por lo que no alcanzó a emitir las recomendaciones correspondientes.

Previa exposición de la problemática detectada, las recomendaciones a las políticas evaluadas en 2010 se dividieron en grupos, manteniendo uniformidad en su formato al incluir el objetivo de la misma, la dependencia responsable y el plazo para cumplirse, además de seguir una secuencia en su presentación. Debido a ello presentan una notable claridad y coherencia, aunque en algunos casos, su extensión y diversidad les puede restar especificidad. En general, éstas reflejan el trabajo y las recomendaciones emitidas por los distintos equipos evaluadores.

La gran mayoría de las recomendaciones se dirigieron principalmente a aspectos de diseño, relacionados con la introducción de temas en programas o incluso con la elaboración de uno nuevo, algunas más se refirieron a temas normativos. Las de operación aludieron al reforzamiento de acciones que ya están en marcha. En la política de equidad de género, que tiene más recomendaciones sobre operación, éstas se refieren a cuestiones ligadas con el presupuesto y su correcta aplicación. Tentativamente, en doce de las recomendaciones se percibe una fuerte repercusión presupuestal, aunque la gran mayoría requiere recursos o nuevas partidas para ponerse en marcha (Cuadro 29).

Respuesta institucional a las recomendaciones y difusión general

Sobre la difusión de los resultados de las evaluaciones, el Consejo siempre ha estado a favor de que toda la información sea pública. Una vez aprobado el informe final de las mismas, se convocó a una conferencia de prensa donde se

CUADRO 29
Recomendaciones emitidas por Evalúa DF en 2011-2012

Política	Diseño	Operación	Total
TPCP	26	6	32
Participación ciudadana	5	1	6
Equidad de género	10	9	19
Materno-infantil	0	0	0
Total	41	16	57

Fuente: elaboración propia con base en los informes de recomendaciones emitidos por Evalúa DF a cada una de las políticas evaluadas.

anunciaron las evaluaciones realizadas; después de eso se publicaron íntegros los informes finales en la página del Evalúa DF,¹⁸ también se publicaron notas en prensa. Salvo el informe de evaluación de los servicios de salud materno-infantil, todos los demás se encuentran en la página electrónica del Consejo, al igual que las recomendaciones emitidas.

Respecto a las respuestas por parte de las dependencias a las recomendaciones del Evalúa DF, éstas no siempre manifiestan con claridad su aceptación o rechazo, a veces son ambiguas y se justifican con acciones que estarían en marcha. La negativa de las dependencias para actuar se justifica principalmente con el argumento normativo de que la acción recomendada no es de su competencia o que sus facultades son limitadas. Otras razones tienen que ver con la insuficiencia de tiempo para realizar los estudios pertinentes o la falta de recursos para ponerla en marcha. Sobresale la Secretaría de Desarrollo Social (SDSDF) que, al rechazar cuatro de las cinco recomendaciones de la evaluación de la política de participación ciudadana, solicita una prórroga para acordar con Evalúa DF la interpretación y alcances de las mismas.

Existen dependencias que aceptan parcialmente las recomendaciones debido a que ya existen acciones al respecto que contribuyen a su realización, y en

¹⁸ Entrevista al maestro Yánes Rizo, ya mencionada.

algunos casos sólo se comprometen a darles seguimiento y/o fortalecerlas. Las dependencias que aceptan las recomendaciones de manera total, como Inmujeres, condicionan su cumplimiento a la asignación o ampliación del presupuesto, estableciendo en algunos casos una fecha para recibir los recursos y poder cumplir en el plazo previsto; otras dependencias, al tratarse de una acción interinstitucional, muestran su disponibilidad para colaborar y coordinarse con los demás involucrados.

Un caso especial lo constituye la Secretaría de Transportes y Vialidad (Setravi), responsable de 26 recomendaciones que considera pertinentes pero que requieren mucho más tiempo y presupuesto para cumplirse, por lo que en general acepta iniciar solicitudes, gestiones, diseñar propuestas, impulsar estrategias o realizar avances para su cumplimiento. En el Cuadro 30 se presenta una aproximación al número de recomendaciones aceptadas y rechazadas por parte de las dependencias. La aproximación se debe a que no todas las dependencias se pronunciaron ante las recomendaciones del Evalúa DF¹⁹ y cuando lo hacen, algunas no declaran abiertamente su aceptación o rechazo (principalmente en el caso de TPCP), por lo que hubo que interpretar su respuesta.

¹⁹ O en su caso, no estuvo disponible el documento en el sitio web y no se logró obtener respuesta del Evalúa DF ante la solicitud de información al respecto.

CUADRO 30
Respuesta de las dependencias a las recomendaciones

Dependencia	Recomendaciones	Aceptadas	Aceptadas parcialmente	No aceptadas
TRANSPORTE PÚBLICO COLECTIVO DE PASAJEROS				
Setravi	26	20		6
Secretaría de Finanzas	2			2
Servicio de Transporte Eléctrico	8	8		
RTP**	9	1		8
Sistema de Transporte Colectivo**	12	1	8	3
Seduvi**	3	1		2
COPRED****	2	2		
Inmujeres****	2	2		
PGJDF****	1	1		
Secretaría de Desarrollo Social***	2	2		
Secretaría de Medio Ambiente***	1	1		
Total	68	39	8	21
PARTICIPACIÓN CIUDADANA				
Secretaría de Desarrollo Social	5	1		4
Inmujeres-DF	1			1
Total	6	1		5
EQUIDAD DE GÉNERO				
Inmujeres-DF	12	5	7	
Secretaría de Finanzas	5		2	3
Secretaría de Salud	2		2	
Secretaría de Desarrollo Social	1****			1
Oficialía Mayor	1****			1
Total	21	5	11	5
Total de las tres políticas	95	45	19	31

* La gran mayoría de las recomendaciones son interinstitucionales, por lo que existen más dependencias involucradas, pero no se cuenta con una respuesta de su parte.

** No hacen explícita su aceptación o rechazo.

*** Coadyuvan junto con otras instituciones.

**** Comparten la recomendación con Inmujeres-DF.

Fuente: elaboración propia con base en los informes de recomendaciones emitidos por Evalúa DF a cada una de las políticas evaluadas, y los documentos de respuesta emitidos por las dependencias ante dichas recomendaciones.

Desempeño del Programa de Cambio Climático de la Ciudad de México 2008-2012

Contexto de la evaluación

Como se señaló en el capítulo 11, desde inicios de la década pasada el gobierno de la Ciudad de México ha puesto en marcha acciones para enfrentar el cambio climático. En 2004 diseñó su Estrategia Local de Acción Climática y recientemente su Estrategia de Acción Climática para el periodo 2014-2020. También ha ejecutado el PACCM 2008 a 2012 y ahora ha iniciado el correspondiente a 2014 a 2020. En el 2011 publicó la Ley de Mitigación y Adaptación al Cambio Climático y Desarrollo Sustentable para el Distrito Federal.

Hasta 2011 no contaba con evaluaciones que permitieran identificar y valorar los efectos de sus acciones, sólo existían algunos estudios realizados por el propio GDF para informar los avances de sus acciones y proyectos. En 2012, la Sedema solicitó la evaluación del PACCM 2008-2012 que si bien en el tercer párrafo de la introducción afirma que el objetivo de la misma es “[...] realizar una revisión del desempeño de los proyectos y acciones que integran el PACCM y generar recomendaciones que permitan fortalecer la contribución de la Ciudad de México a la reducción de emisiones de GEI, así como mejorar su resiliencia a los impactos del cambio climático” (CMM, 2012:14); lo cierto es que el documento de la evaluación presenta estudios que contribuyen a realizar de manera parcial una evaluación del diseño y operación del Programa de Acción Climática de la Ciudad de México.

Si bien cada organismo encargado de algún proyecto del PACCM 2008-2012 es responsable de presentar anualmente ante la Sedema los avances y logros obtenidos respecto a los objetivos planteados en el mismo, para contar con un respaldo de la credibilidad de la información y atender las debilidades que se detecten, como se declara en los TR, la Sedema solicitó la evaluación.

Marco normativo

La solicitud mencionada se sustenta en la Ley de Mitigación y Adaptación al Cambio Climático y Desarrollo Sustentable del Distrito Federal, publicada el

16 de junio de 2011. Además, como parte de los compromisos del gobierno federal plasmados en la ENCC, le corresponde a la Sedema aplicar el MRV al Programa de Acción Climática de la Ciudad de México así como a los Proyectos de Acción Climática Subsecuentes (artículo 8, fracción II).

La Ley también establece que la Comisión Interinstitucional de Cambio Climático del Distrito Federal (CICCDF) es la instancia responsable de la coordinación, evaluación, medición, verificación y revisión del Programa de Acción Climática de la Ciudad de México (artículo 3, fracción IV); sin embargo en los TR no figura dicha Comisión. Sólo presenta a la Dirección del Programa de Cambio Climático y Proyectos MDL,²⁰ adscrita a la Sedema, como una de las responsables administrativas y técnicas del proyecto Evaluación del PACCM, la que funge como colaboradora de la Secretaría Técnica de la Comisión.

Para guiar la evaluación del PACCM se elaboraron los TR del proyecto “Evaluación del Programa de Acción Climática de la Ciudad de México 2008-2012”, fechados el 19 de junio de 2012. En este documento se especificó desde el inicio que dicho estudio busca realizar una revisión del desempeño de los proyectos y acciones que integran el PACCM para los periodos 2008-2012, aunque en los resultados esperados se dice que se espera una evaluación de dicho desempeño y recomendaciones, tanto de las metodologías y los indicadores aplicados, como de la infraestructura y los recursos humanos requeridos para el cumplimiento y la adecuada evaluación de objetivos y metas del Programa de Acción Climática de la Ciudad de México.

Para alcanzar su objetivo los TR plantean los siguientes objetivos específicos: 1) elaborar un reporte de revisión del cumplimiento de los proyectos y acciones; 2) estimar el nivel de reducción de emisiones de gases de efecto invernadero; 3) evaluar el éxito de implementación de las medidas y estimar el grado de avance; 4) estimar un reporte de la evaluación de las estimaciones de las reducciones de emisiones de gases de efecto invernadero; 5) hacer un análisis de los requerimientos para cumplir con las exigencias de la metodología del Panel Intergubernamental sobre Cambio Climático (IPCC, por sus siglas en inglés); 6) presentar recomendaciones para mejorar las metodologías utilizadas para calcular las reducciones de emisiones en las acciones; 7) presentar sugerencias

²⁰ MDL significa mecanismo de desarrollo limpio.

para mejorar la trazabilidad de la información y datos utilizados para el cálculo de las metas; 8) hacer un reporte detallando las metodologías de evaluación empleadas; 9) proponer las metodologías para realizar evaluaciones de manera periódica; y 10) elaborar y publicar una guía metodológica para la evaluación del Programa de Acción Climática de la Ciudad de México.

De igual forma, plantean que la evaluación tendría que ser desarrollada en tres fases: planeación de los trabajos, ejecución de la verificación y presentación de los resultados, en un periodo máximo de seis meses. Además se especifican las actividades a desarrollar y productos que se esperan. El director general de Planeación y Coordinación de Políticas de la Secretaría de Medio Ambiente del GDF y el director del programa de Cambio Climático y Proyectos MDL fueron los responsables administrativos y técnicos del proyecto de evaluación del Programa de Acción Climática de la Ciudad de México.

Equipo evaluador y recursos económicos

La contratación se realizó por adjudicación directa, sin mediar un proceso de convocatoria y selección entre candidatos, al CMM y el precio de la evaluación fue de \$5 200 000.00. Este centro constituye una asociación civil creada en 2004 y enfocada a la investigación y promoción de políticas públicas; realiza principalmente estudios estratégicos sobre energía y medio ambiente, cambio climático y calidad del aire.

El equipo evaluador estuvo conformado por 17 integrantes, en su mayoría con licenciatura en ingeniería, biología y química farmacobióloga; sólo tres de ellos cuentan con una maestría. La líder del proyecto es licenciada en economía y ocupa el puesto de economista de proyectos dentro del CMM. La principal debilidad que se observa del equipo evaluador es la carencia de colaboradores con una mayor formación académica; así como la ausencia de un experto en evaluación de proyectos ambientales o en políticas públicas. En los TR no se especificó el perfil requerido para el equipo evaluador.

Metodología de evaluación

El documento entregado por el CMM carece de un apartado específico donde se presenten los métodos y técnicas de investigación, así como las fuentes de información utilizadas para realizar los tres estudios que fueron convenidos, por lo que este aspecto se analiza en cada uno de los informes entregados.

El primero es un estudio de la instrumentación y seguimiento de las acciones y proyectos del PACCM, que tampoco enuncia su metodología. En cuanto a las fuentes, de acuerdo con lo solicitado en los TR, se hizo uso de información de gabinete, principalmente de documentos oficiales (PACCM 2008-2012, Informes del PACCM 2009-2012), pero carece de información recabada en campo mediante encuestas o entrevistas. Utiliza una guía metodológica, elaborada por el propio CMM, que propone un conjunto de procedimientos y criterios para caracterizar la integridad de la estructura, contenido, gestión y difusión de un programa en materia de cambio climático y que también fue solicitada en los términos de referencia.

El segundo es el estudio de la trazabilidad de la información que sustenta los avances de las acciones del PACCM. A diferencia del estudio anterior, presenta de manera clara y detallada el método y las dos técnicas de investigación utilizadas, la primera referida a una revisión documental (documentos oficiales) y por internet (portales del GDF) y la segunda enfocada a una entrevista a más de 100 funcionarios públicos responsables de la instrumentación y seguimiento de las acciones del PACCM 2008-2012, como fuente primaria. Se informa a quien se entrevistó y la forma de la entrevista (personal, telefónica o mediante correo electrónico). También se detallada el método de análisis que, al igual que las técnicas de investigación utilizadas, es adecuado a lo solicitado en los términos de referencia.

El tercer estudio se refiere a valorar la incertidumbre de las 10 acciones de mitigación contenidas en el PACCM, que son consideradas las más relevantes por su contribución a la mitigación de gases de efecto invernadero. El estudio se realiza a partir de analizar las metodologías empleadas por el GDF para calcular la reducción de emisiones de gases de efecto invernadero. No presenta en un apartado específico el método de recolección de información seguido, tampoco hace referencia a las técnicas de investigación y el método de análisis, así como su pertinencia y justificación. Destaca que el CMM solicitó de manera formal

a la Sedema información sobre el inventario de emisiones de gases de efecto invernadero, la cual tuvo que ser reelaborada a partir de ciertas indicaciones del equipo evaluador. Esta petición se debió a que, a juicio del grupo, no se contó desde un inicio con un inventario de emisiones de gases de efecto invernadero consolidado, que permitiera el establecimiento de una línea base para su periodo de aplicación 2008-2012.

Los métodos y técnicas utilizados para la realización de los tres estudios se ajustan a lo solicitado en los términos de referencia.

Informe final

El informe final de la evaluación del programa de 216 páginas (sin anexos) está compuesto por los siguientes seis apartados: 1) marco institucional, 2) evolución y cumplimiento del PACCM, 3) evaluación de la trazabilidad de la información, 4) evaluación de metodologías empleadas para estimar reducción de emisiones, 5) metodología propuesta para el seguimiento del PACCM, y 6) observaciones y recomendaciones a las acciones del PACCM. Incluye también una introducción, un apartado de referencias y cuatro anexos: el primero referido a la agenda de entrevistas realizadas a funcionarios responsables de las medidas del PACCM, el segundo a la evaluación de trazabilidad de cada una de las medidas, el tercero presenta las cédulas de evaluación de trazabilidad del programa, y el último ofrece los criterios para evaluar un Programa de Acción Climática.

• Introducción

El informe inicia con una introducción en que se presenta brevemente el papel que ha tenido el GDF en la gestión del cambio climático, en particular con el PACCM 2008-2012, y destaca que, con la evaluación efectuada, la Ciudad de México se posiciona a la cabeza de América Latina en la realización de un trabajo de este tipo.

En la misma se informa el objetivo de la evaluación y se anuncian los tres estudios que se habrían realizado para alcanzarlo: 1) revisión de la evolución de las distintas acciones y proyectos reportados en el PACCM; 2) evalua-

ción de la trazabilidad de la información que sustenta los avances; y 3) evaluación de la incertidumbre de las 10 acciones de mitigación más relevantes por su contribución a la reducción de gases de efecto invernadero. Tanto el objetivo como los estudios son presentados de manera clara, más no precisa y cumplen con la mayoría de lo solicitado en los términos de referencia.²¹

Puesto que una evaluación de programas públicos busca valorar su diseño, operación, resultados o impacto con el fin de mejorar el funcionamiento del mismo y obtener mejores resultados para la población beneficiaria,²² los dos últimos estudios no pueden ser considerados “evaluaciones” a pesar de que en el informe se les tipifica como tales,²³ sino investigaciones que producen conocimiento necesario para evaluar los resultados de la ejecución del PACCM 2008-2012. Por otra parte, las evaluaciones requieren obtener conclusiones sobre los problemas identificados pero también proponer recomendaciones para reducirlos o eliminarlos. Sin embargo, en la revisión de la evolución de las acciones del PACCM faltan estas últimas,²⁴ lo que muestra que incluso este primer estudio, el más cercano a una evaluación de programas, tampoco cumple plenamente con todos los contenidos que este tipo de trabajos implica.

Por último, la introducción menciona que, junto a los tres estudios referidos, se presentan otros dos productos: el primero es una herramienta informática orientada a mejorar la calidad de la información reportada y la periodicidad

²¹ Uno de los objetivos específicos planteados en los TR es “evaluar el éxito de implementación de las medidas de mitigación y estimar el grado de avance de las mismas”; para lograrlo se solicita considerar dos elementos: el primero, la metodología utilizada para calcular el nivel de cumplimiento respecto a la meta y por lo tanto los avances de las acciones establecidas en el programa; el segundo, es la trazabilidad de la información y datos utilizados para calcular los avances de las acciones. Efectivamente los dos últimos estudios atienden estos elementos, sin embargo sólo se consideran para realizar la valoración del éxito de implementación.

²² En el caso del PACCM la población beneficiaria está representada por los 21 millones de habitantes de la Zona Metropolitana de la Ciudad de México, que incluye los 8.7 millones de habitantes del Distrito Federal.

²³ Aunque estrictamente el título del primero sólo dice “Revisión” y no evaluación.

²⁴ En el caso de la trazabilidad de la información se incluyen tanto conclusiones como recomendaciones y en el correspondiente a la incertidumbre de las acciones de mitigación no se presentan ninguna de las dos.

con la cual se realizan las mediciones de los avances y el segundo es una “Guía metodológica para la evaluación de programas de acción climática de gobiernos locales”, los cuales también fueron solicitados en los TR. También informa que el financiamiento de los trabajos se realizó con recursos del Ramo 16 (medio ambiente) del PEF; sin mencionar el monto asignado.

La introducción presentada es pertinente, no así clara y suficiente para ser una primera aproximación al contenido del trabajo, ya que no se expone la importancia que tiene la evaluación de los programas de acción climática, los retos o dificultades que enfrentan aquellos que deciden hacerlo, no se destaca su apego a lo solicitado por la Sedema, no se menciona la dependencia gubernamental que debería dar seguimiento al programa y a la propia evaluación. El abordaje de estos temas hubiese permitido explicar, por ejemplo, la razón por la que sólo se requirió una revisión del desempeño de los proyectos y acciones del PACCM 2008-2012 en lugar de solicitar una evaluación de su impacto. De igual forma no se mencionan los métodos y técnicas de investigación utilizados en cada estudio.

Si bien en ella no se presenta el marco institucional en el que se sustenta el PACCM 2008-2012, éste se describe en su primer capítulo. Dicha descripción inicia muy acertadamente con el Programa General de Desarrollo del Distrito Federal 2007-2012 para después presentar todos aquellos instrumentos de planeación que se han construido para atender el problema del cambio climático. También presenta el objetivo de cada instrumento, resaltando su relación con el cambio climático y en particular, las acciones que propone en torno a esta problemática.

- Marco teórico-conceptual

El informe no presenta un apartado de marco teórico-conceptual general para realizar los tres estudios ya mencionados; sin embargo, se incluyen marcos específicos en dos de éstos: trazabilidad de la información e incertidumbre de las 10 principales acciones de mitigación. En el primero de éstos se presenta un apartado breve sobre el concepto de trazabilidad, en el marco de un Programa Estatal de Acción Climática, aunque este abordaje resulta insuficiente debido a la falta de un apartado que trate de manera clara y detallada la evaluación en los programas de acción climática de gobiernos locales y los principales

conceptos, enfoques y criterios que se utilizan para realizarlas. En cambio, en el segundo se incluye un marco referencial amplio que permite tener claridad de los conceptos y procedimientos que se utilizan para estimar la reducción de emisiones. Desafortunadamente, en los TR no se especificaron los aspectos que deberían incluirse dentro de este marco, por lo que el tratamiento del tema resulta insuficiente.

- Calidad de los estudios

Si bien en los TR y en el documento que integra los tres informes de los trabajos realizados no se dice que se realizará una evaluación del diseño y operación del PACCM, lo cierto es que los tres estudios realizados por el equipo evaluador forman parte de una valoración de estos puntos; aunque no son los únicos.

El documento que integra los tres informes de los trabajos realizados no presenta un diagnóstico del problema de cambio climático en el Distrito Federal ni los antecedentes del programa evaluado. De igual manera, no informa sobre las características del programa, los cambios y relaciones con el marco institucional que incidieron en la implementación del mismo, los recursos con los que contó, los factores que impidieron que se llevaran a cabo ciertas acciones, que otras se integraran en fechas posteriores a la instrumentación del PACCM, que se redefinieran sus objetivos o que sólo se lograra cierto nivel de avance. Esta ausencia de antecedentes provoca que el trabajo se limite al seguimiento de avances de una lista de acciones y proyectos, sin medir el nivel de cumplimiento de los compromisos asumidos y mucho menos explicarlo, lo que constituye parte de una evaluación.

El grupo de trabajo tampoco analizó aspectos que en cualquier estudio de este tipo tienen que ser considerados como, por ejemplo, el contexto evaluativo, la claridad de sus objetivos, la población objetivo o la consistencia interna de sus componentes, entre otros. Probablemente, esto se debió a que este aspecto no fue recuperado en los objetivos específicos de los TR, ya presentados.

Lo que sobresale del primer informe es la revisión amplia y adecuada de las acciones realizadas en el marco del PACCM, así como su avance. Aquí se presenta la descripción y conclusiones sobre cada una de las acciones; sin embargo el

estudio se limita a describir, se observa una ausencia de análisis en torno a recursos empleados, consistencia entre operación-diseño, logro de resultados esperados, así como sustentación y retroalimentación de procesos, entre otros; aspectos imprescindibles a considerar para realizar una evaluación del éxito de la implementación de las medidas de mitigación.

- Conclusiones y recomendaciones

Los tres informes que integran la evaluación del PACCM 2008-2012 recuperan los hallazgos que fueron identificados a lo largo de los análisis y dedican un apartado en que se exponen los resultados alcanzados, así como sus valoraciones y conclusiones.

Los evaluadores emiten juicios valorativos de los tres aspectos evaluados; sin embargo, éstos resultan insuficientemente fundamentados en el trabajo, debido a las debilidades mostradas en los apartados anteriores. Concretamente, el primer informe presenta una conclusión referida a la valoración de la instrumentación y el avance de las acciones del programa, solicitada en los TR, pero es difícil de sostener puesto que no se deduce del estudio realizado, que sólo se limitó a describir el proceso.

A pesar de que las recomendaciones no se presentan claramente al final de cada estudio, éstas se exponen en el último apartado del documento. Las recomendaciones son amplias, fundamentadas, puntuales y acordes con lo requerido en los TR y se acompañan por lo que el equipo evaluador llama “áreas de oportunidad”. Sin embargo, una parte de éstas no se desprende de la problemática analizada en los estudios, especialmente en el primero.

Las recomendaciones giran en torno a continuar con ciertas acciones que son consideradas como área de oportunidad; reconsiderar algunas medidas y expandir o replicar otras. Se acompañan de una explicación breve de los errores en que se incurrió en la instrumentación de la acción o bien en la medición de los avances, lo que tendría que haberse desarrollado ampliamente en el propio estudio.

También se propone realizar una clasificación de la actividad del programa en acciones clave y coadyuvantes,²⁵ crear instancias²⁶ encargadas de calcular las emisiones mitigadas, así como de fijar y supervisar la aplicación adecuada de las metodologías en el sector transporte y “refinar la estrategia climática del Gobierno del Distrito Federal” (CMM, 2012:210), depurando algunas medidas y diseñando otras dentro del próximo PACCM, mismas que no se acompañan de una argumentación que las fundamente.

Al término del último apartado se presenta otro grupo de recomendaciones que fueron requeridas de manera puntual en los TR, destinadas a mejorar la trazabilidad de la información, la infraestructura física y humana, la metodología para reportar la reducción de emisiones y la revisión de las mejores prácticas para el próximo periodo del PACCM; sin embargo, no todos estos aspectos fueron analizados en los estudios.

En síntesis, las conclusiones y recomendaciones del equipo evaluador resultan parcialmente adecuadas y poco fundamentadas en los trabajos.

Valoración de la evaluación

El equipo de trabajo del CMM cumplió con la mayor parte de lo solicitado en los TR, sus estudios representan un importante esfuerzo por proporcionar información que contribuya a mejorar la actuación del GDF en la atención del problema del cambio climático. Sin embargo, en futuras evaluaciones se deben atender las debilidades evidenciadas en este trabajo, es el caso de exponer la importancia que tiene la evaluación de los programas de acción climática, los retos o dificultades que enfrentan aquellos que deciden hacerlo, el marco jurídico en el que se sustenta este tipo de evaluaciones, el diseño de un marco teórico-conceptual sólido; la consideración de todos los aspectos a valorar en evaluaciones de este tipo, así como la evaluación del éxito de la implementación de las medidas, esto último solicitado expresamente en los términos de referencia.

²⁵ Debido a que la cantidad de acciones inicialmente planteadas y progresivamente añadidas al PACCM complicó su instrumentación, seguimiento y sobre todo su evaluación.

²⁶ Comisión técnica de compuestos de efecto invernadero y grupo de trabajo para el sector transporte.

La evaluación del PACCM es una actividad necesaria para mejorar su funcionamiento, pero insuficiente para valorar los logros alcanzados y valorar sus efectos en las condiciones de vida de la población beneficiaria. Por lo que resulta indispensable realizar una evaluación de impacto del programa que permita valorar los logros alcanzados en comparación con sus efectos esperados y saber si ha contribuido a la atención del problema que justifica la acción pública.

Por último, las evaluaciones y sus resultados no sólo deben ser accesibles a quienes deciden y ejecutan las políticas sino también deben ser transparentes para todas las partes interesadas, incluyendo a la ciudadanía en general.

Difusión del informe final y respuesta institucional a la evaluación

Hasta el momento, los resultados del Proyecto “Evaluación del Programa de Acción Climática de la Ciudad de México 2008-2012” se han compartido con funcionarios de las 16 delegaciones políticas del Distrito Federal, con personal de la Dirección del Programa de Cambio Climático y Proyectos MDL, con colaboradores de la Secretaría Técnica de la CICCDF y con la población en general a partir de congresos, foros, seminarios, conferencias, cursos y talleres.

El resumen de la evaluación realizada también se ha difundido a través de la página web de la Sedema y del CMM,²⁷ pero no el documento completo donde se presentan los estudios realizados en el marco del proyecto y sus términos de referencia. En síntesis, se carece de un acceso público a los resultados de esta evaluación, elemento necesario para lograr la transparencia y rendición de cuentas por parte de la actividad gubernamental.

A la fecha no existe una reacción formal ante la evaluación por parte de la Dirección General de Planeación y Coordinación de Políticas de la Sedema y de la Dirección del Programa de Cambio Climático y Proyectos MDL del GDF, responsables del proyecto. Sólo se realizó una invitación al CMM por parte de la última dirección mencionada y de la CICCDF para informar de los resultados de

²⁷ Información proporcionada por el encargado de la Oficina de Información Pública de la Sedema del Distrito Federal, con fecha 29 de octubre de 2014.

la evaluación solicitada a finales de 2012. Sin embargo, para la elaboración del Programa de Acción Climática de Ciudad de México 2014-2020 se tomaron en consideración todas sus sugerencias.

Conclusiones y recomendaciones parciales

Las cinco evaluaciones revisadas en este apartado pertenecen al GDF pero presentan características muy diferentes entre sí porque las primeras cuatro, aunque varían sus alcances y temas, pertenecen todas a un mismo sistema (Evalúa DF), en cambio la quinta es independiente del anterior (Sedema). Todas contaron con TR específicos a cada caso.

Los equipos evaluadores fueron interdisciplinarios, la mayoría de buen nivel académico, adecuados al tema, pero sólo algunos habían tenido experiencias previas en trabajos de evaluación. Las convocatorias del Evalúa DF publican los montos disponibles para cada trabajo: \$348 000.00 para los estudios de gabinete en 2010 y \$1 050 000.00 para el que incluyó trabajo de campo en 2011. Los equipos contratados fueron los ganadores dentro de una convocatoria abierta entre quienes cumplieron los requisitos y presentaron la mejor propuesta técnica, con base en criterios preestablecidos. En cambio, el quinto trabajo se contrató en 2012 por adjudicación directa a un centro de elevado prestigio y se pagaron \$5 200 000.00 por su realización, siendo el tercero más caro (después de las evaluaciones de impacto del PDHO y PEI) de los que se tuvo información en este libro. Aunque la comparación no es fácil entre evaluaciones tan diferentes, considerando la complejidad de lo solicitado en cada uno de los TR y los resultados entregados; tampoco es evidente la justificación del último monto frente a los estudios de gabinete y trabajo de campo mucho más económicos de las cuatro evaluaciones previas.

Los trabajos contratados por Evalúa DF evaluaron políticas y servicios públicos y utilizaron fundamentalmente entrevistas y análisis de documentos como métodos de investigación, enriquecidos con encuestas, observación *in situ*, grupos focales y análisis estadístico en el caso de la evaluación de impacto; la contratada por la Sedema evaluó un programa público y empleó entrevistas a funcionarios públicos, análisis documental y otros métodos de carácter técnico

no expresamente señalados en su informe. Estos son suficientes en general y adecuados a los solicitados por los TR, salvo en el caso de la evaluación de impacto que, aunque muy útil, no logró responder algunas preguntas y analizar la posible influencia de variables extrañas en los efectos atribuidos a los servicios evaluados.

Los informes fueron extensos (entre 186 y 395 páginas) y, en general, se presentan adecuadamente estructurados y redactados, aunque subsisten algunos aspectos confusos; los del Evalúa DF están disponibles en su página electrónica donde se incluye además la respuesta institucional ante la evaluación y sus recomendaciones, no así el de Sedema que ha sido un proceso mucho menos transparente.

Se incluyen trabajos aceptables y otros de buena calidad, de difícil comparación por la mayor complejidad técnica del caso de transporte y especialmente en materia de cambio climático. Como consecuencia de lo anterior, las recomendaciones propuestas para los problemas identificados son también variadas.

En los casos de las evaluaciones contratadas por Evalúa DF: *a*) disponer de más tiempo para la realización de los trabajos; *b*) no solicitar contenidos que rebasan el objetivo de la convocatoria; *c*) exigir mayor experiencia en evaluación a los equipos contratados; *d*) presentar expresamente los métodos de investigación utilizados para analizar la información; *e*) identificar claramente en todos los casos la relación entre datos, análisis, conclusiones, valoraciones y recomendaciones; *f*) especificar la instancia responsable y el plazo para cumplir las recomendaciones; *g*) acompañar siempre el informe final de un resumen ejecutivo; y *h*) dar seguimiento continuo a las recomendaciones aceptadas y proceder a iniciar las controversias correspondientes en los casos de rechazo.

En la evaluación realizada por el CMM: *a*) utilizar con precisión el concepto de evaluación en los TR, diferenciándolo de otros cercanos (revisión, estudio, análisis, etcétera); *b*) expresar con claridad y consistencia el objetivo general y los específicos de la evaluación, así como los productos implicados en los TR y recuperarlos en los mismos términos en el informe del trabajo; *c*) indicar en los TR el plazo para la realización del trabajo y en el informe el tiempo efectivamente utilizado; *d*) señalar el marco jurídico en que se sustenta la evaluación; *e*) efectuar convocatorias más abiertas y hacer públicos los TR; *f*) integrar en el equipo de trabajo especialistas en evaluación de programas; *g*) incluir un marco teórico-conceptual en el informe; *h*) presentar claramente los distintos

métodos de evaluación utilizados; *i*) revisar todos los aspectos básicos de una evaluación de diseño y operación de programas; *j*) presentar ordenadamente las recomendaciones y fundamentarlas en el análisis efectuado; *k*) ampliar la publicidad y el acceso al informe final de evaluación; y *l*) dar seguimiento a sus recomendaciones.

Conclusiones y recomendaciones generales de la segunda parte

Myriam Cardozo Brum

Ante todo es importante recuperar lo expresado en el primer capítulo de esta segunda parte en el sentido de que, a nivel federal, se hizo la metaevaluación de una diversidad de sectores y tipos de evaluación distribuidas en el periodo 2007-2014 pero esto de ninguna manera garantiza que sean representativas de un conjunto que ha girado en torno a las 2 000 evaluaciones anuales. De serlo, nuestra muestra indicaría un relativo estancamiento de la calidad en niveles mayoritariamente aceptables y sin tendencia a mejorar, pero lo que es innegable es que, salvo 2007-2008 en que sólo se analizó una evaluación muy buena, en todos los años restantes aparece siempre alguna de baja calidad o apenas aceptable. Esto permite un acercamiento a la calidad de los trabajos en la medida que encontramos ejemplos que muestran tanto que muy buenas evaluaciones son posibles (básicamente tres en nuestra muestra de 16) como otras (seis en nuestra muestra) que no realizan aportación alguna.

Con la intención de sintetizar el trabajo de los capítulos previos para facilitar la presentación de las conclusiones, en el Cuadro 31 se resumen los principales resultados de las metaevaluaciones federales realizadas, organizados por sectores gubernamentales.

Es importante resaltar que en nueve de los 16 casos considerados no se localizaron sus TR y los siete restantes fueron del tipo genérico aplicado a todas las evaluaciones de diagnóstico (estrictamente lineamientos), diseño y las EED. En seis casos no existe información alguna sobre la forma de contratación del equipo que haría el trabajo y sólo en dos se menciona una convocatoria por

CUADRO 31

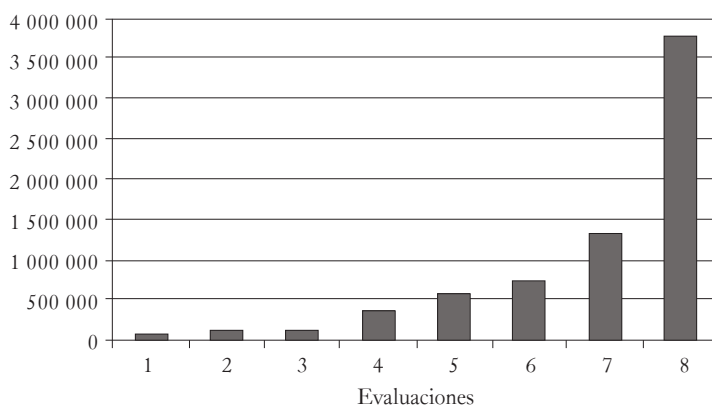
Principales resultados de las metaevaluaciones federales

Temas	Desarrollo social	Educación	Medio ambiente	Salud	Desarrollo rural
Tipo de evaluación	1 procesos 1 integral desempeño 1 estratégica 2 impacto	1 diagnóstico 1 específica 1 EFD 1 impacto	2 diseño 1 indicadores	1 ECR 1 complementaria	2 ECR
Existencia TR	1	1	2	1	2
Calificación equipo	5 alta y muy alta, composición no <i>ad hoc</i>	1 interna 1 aceptable 2 s/información	1 interna 2 aceptables	2 aceptables s/ información completa	2 alta
Forma contratación	2 asignación directa 3 s/información	1 interna 1 invitación a 3 2 s/información	1 interna 2 asignación directa	1 invitación a 3 1 s/información	1 convenio 1 asignación directa
Costos	\$1 300 000.00 \$34 839 000.00 3 s/información	1 interna s/costo \$50 496.00 2 s/información	1 interna s/costo \$ 92 220.00 \$102 080.00	\$569 931.00 \$3 758 344.00	\$347 725.00 \$730 000.00
Métodos investigación	Análisis documental Encuestas Diseño experimental	Análisis documental Cuestionarios Diseño experimental	Análisis documental Cuestionarios Reuniones de trabajo	Análisis documental Cuestionarios Reuniones de trabajo Observación directa	Análisis documental Entrevistas
Tamaño informe	40 a 315 pp.	2 a 38 pp.	33-38 pp.	63-154 pp.	174-182 pp.
Valoración calidad	2 aceptables 2 buenas 1 muy buena	4 no aceptables	2 no aceptables 1 muy buena	1 aceptable 1 muy buena	2 buenas

Fuente: elaboración propia con la información de los capítulos de la segunda parte.

invitación a cuando menos tres instituciones. Se debe destacar que, a partir de 2014, el Inventario de Programas y Acciones Federales de Desarrollo Social ya no publica dicha información, reduciendo la transparencia en la materia. En la mitad de las evaluaciones revisadas se conoce su costo que va desde \$50 496.00 por una EED hasta \$34 839 000.00 por una de impacto (Gráfica 1 donde los últimos dos datos no se alcanzan a graficar debido a su magnitud).¹

GRÁFICA 1
Costos de las evaluaciones federales



1. EED de Escuelas de Tiempo Completo.
2. Diseño de Estabilización de Cuencas y Acuíferos.
3. Diseño del Promanp.
4. ECR de Apoyo al Ingreso Agropecuario Procampo.
5. ECR Seguro Popular.
6. ECR de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural.
7. Procesos de Fomento y Desarrollo de Culturas Indígenas.
8. Complementaria de Sicalidad.

Fuente: elaboración propia con datos de los capítulos precedentes.

¹ Se trata de \$34 839 000.00 y \$10 000 000.00 correspondientes a las evaluaciones de impacto del Programa de Desarrollo Humano Oportunidades y del Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras.

Sería esperable que los costos se correlacionen con la calidad de los trabajos (Cuadro 32), lo que se cumple en la mayoría de los casos con dos excepciones opuestas: una evaluación de diseño de bajo costo y muy alta calidad y una complementaria de alto costo y calidad meramente aceptable.

Continuando con el análisis de la información del Cuadro 31, se observa que los equipos evaluadores con calificaciones más altas se ubican en Desarrollo Social (aunque su composición interdisciplinaria no es la óptima), seguida de Desarrollo Rural, mientras dos casos de Educación carecen de toda información al respecto. Los métodos de investigación son variados aunque todas las evaluaciones coinciden en utilizar el análisis documental, especialmente de origen institucional. Pocos estudios realizan trabajos de campo, pero un aspecto muy descuidado en estos casos es la exposición de los métodos de análisis de la información recogida y, cuando se realizan encuestas por muestreo, la determinación del tamaño muestral y la forma de seleccionar a las unidades integrantes de la misma para valorar su representatividad. Tres de los trabajos revisados aplicaron diseños experimentales con resultados menos contundentes de lo esperado por los equipos evaluadores.

Los informes tuvieron dimensiones muy variables, desde dos páginas en el caso de una ficha de monitoreo en educación hasta 315 en una evaluación de impacto de desarrollo social. Este último sector destaca también por no incluir ninguna evaluación de baja calidad y, en cambio, cuenta con una de las tres mejor calificadas. En el extremo opuesto, todas las de educación tienen calidad inadecuada. Uno de los problemas que suele afectarla es la falta de fundamentación de sus conclusiones y propuestas en el trabajo efectivamente analizado.

Si comparamos esta situación con la vivida en 2001-2006 durante la contratación de evaluaciones por la Sedesol (Cardozo 2006), no hay dudas 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 de monitoreo de un conjunto de indicadores elaborado por un agente externo pero sólo con información

CONCLUSIONES DE LA SEGUNDA PARTE

CUADRO 32
Costo y calidad de las evaluaciones federales

Costos (\$)	Calidad no aceptable	Calidad aceptable	Calidad alta	Calidad muy alta
50 496.00	Específica de Escuelas de Tiempo Completo			
92 220.00				Diseño de Estabilización de Cuencas y Acuíferos
102 080.00	Diseño del Promanp			
347 725.00			ECR de Apoyo al Ingr. Agropecuario Procampo	
569 931.00				ECR Seguro Popular
730 000.00			ECR Des. de Cap. Innovación Tecn. y Ext. Rural	
1 300 000.00			Procesos de Fom. y Desarrollo de Culturas Indígenas	
3 758 344.00		Complementaria de Sicalidad		
10 000 000.00			Impacto del PEI	
34 839 000.00			Impacto del PDHO	

Fuente: elaboración propia con datos de los capítulos precedentes.

institucional, que no puede verificar. Otras son documento preparados a partir de evaluaciones previas (i.e. fichas de monitoreo, “metaevaluaciones”). Incluso la extensión del informe era notablemente mayor en el periodo 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) y se acompaña de un importante nivel de falta de información, especialmente en torno a los costos y mecanismos de contratación de trabajos, como se muestra en el Cuadro 31.

La práctica 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 y en el caso particular de desarrollo social.

De las evaluaciones revisadas en nuestra muestra 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.

Por último, es importante recordar que la calidad de las evaluaciones no depende sólo de la normatividad, la información disponible o la metodología aplicada, sino también del enfoque teórico adoptado y la organización del trabajo (Cardozo, 2006). Algunos de estos temas fueron planteados en las entrevistas a funcionarios realizadas por Székely (2013), como el limitado involucramiento de los responsables a cargo del diseño y la operación de los programas en la definición de la estrategia de evaluación, los criterios para especificar el contenido del PAE, el desconocimiento de los esfuerzos de evaluación en programas complementarios, el desfase entre la generación de resultados y los tiempos para la toma de decisiones presupuestales, entre otros.

En cuanto a las recomendaciones más generales a nivel federal, éstas incluyen: *a*) cumplir con la obligación legal de realizar siempre evaluaciones externas; *b*) emitir convocatorias, dirigidas al menos a tres instituciones en todos los casos; *c*) explicitar los criterios de selección de propuestas; *d*) emitir siempre TR para cada tipo de evaluación, adecuarlos a los problemas concretos del programa a evaluar y publicarlos en la páginas web institucionales; *e*) conformar equipos interdisciplinarios de evaluación, experimentados y *ad hoc* a cada programa; *f*) informar siempre los montos contratados; *g*) exponer la metodología utilizada e incluir la realización de trabajos de campo que verifiquen y complementen la información institucional; *h*) exigir un mínimo de 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 transformadas en ASM y su cumplimiento, desglosadas para cada evaluación realizada.

En el caso del GDF no existen problemas con la selección de los trabajos porque se incluyeron todas las evaluaciones contratadas por Evalúa DF en 2010 y 2011, además de otra solicitada por la Sedema. Sus conclusiones y recomendaciones ya fueron presentadas en el último capítulo y desglosadas para ambos casos. 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 sólo programas). Los trabajos también son más amplios (186 a 395 pp.) e intentan profundizar en la explicación de los hallazgos reportados; sólo en algunos casos federales (Escuelas de Tiempo Completo, Diseño de Estabilización de Cuencas y Acuíferos y Promanp) este resultado es explicable por su menor disponibilidad de recursos. En síntesis, la calidad de las evaluaciones de Evalúa DF se califica como aceptable o buena, según los casos. En consecuencia, sus recomendaciones coinciden con las federales en torno a precisar métodos de investigación, fundamentar mejor las conclusiones, especificar más las recomendaciones y darles seguimiento, y agregan asuntos específicos como disponer más tiempo para los trabajos. Finalmente, el trabajo realizado en el PACCM se considera de calidad aunque no haya cumplido estrictamente con el contenido esperado en una evaluación.

En la tercera parte de este libro se analizará si las recomendaciones efectuadas en los trabajos analizados se vuelven compromisos de mejoría y si efectivamente se cumplen, así como si las posibles explicaciones de esa situación guardan relación con la calidad de los trabajos.

TERCERA PARTE
Utilización de las evaluaciones

Introducción

Una de las mayores preocupaciones mundiales en materia de evaluación es que sus resultados sean efectivamente utilizados para mejorar o cancelar programas y ajustar sus recursos. Muchas de las investigaciones y metaevaluaciones de países diversos, especialmente en Estados Unidos de América, señalan que la utilización instrumental, realizada por las propias políticas y programas evaluados, es muy reducida.

Dentro del contexto mexicano, en la parte II del libro se mencionó la percepción de académicos y funcionarios internacionales sobre la necesidad de que Coneval logre que las evaluaciones tengan un mayor impacto en mejorar los programas sociales. Sin embargo, el trabajo revisado no avanzó suficientemente en analizar la utilización de la información que Coneval ofrece porque consideró que no se disponía de los datos necesarios para realizar su valoración completa (Székely, 2013).

En esta tercera parte se analizan, a nivel teórico, los criterios y factores que afectan a los diferentes tipos de utilización posible, y a nivel empírico, se describe la experiencia mexicana y se refiere a otros casos nacionales que pueden dejar enseñanzas sobre cómo obtener mejores resultados.

CAPÍTULO XXI

Revisión de la literatura

Myriam Cardozo Brum

La utilización de las evaluaciones puede realizarse de muchas maneras y no hay una tipología universalmente aceptada. Diversos investigadores (Rich, 1977; Pelz, 1978; Young y Comptois, 1979, citados en Leviton y Hughes, 1981) han identificado tres tipos:

- *Instrumental*, que se presenta cuando hay un caso documentado de un cambio específico del programa como resultado de la evaluación, es decir, que la evaluación está siendo utilizada para la toma de decisiones o resolución de problemas debido a que ha logrado aumentar la certeza de lo que conviene modificar en la política o programa.
- *Conceptual*, que implica influir en el pensamiento o actitud de un diseñador de políticas sobre un tema, poniendo a su disposición información para su posible uso. En este caso, aunque el uso instrumental no ha ocurrido, la información generada se valora positivamente, ha generado un aprendizaje, y podría ser utilizada en el futuro.
- *Persuasivo*, consiste en aumentar la contundencia de un argumento con la intención de convencer a otros para apoyar o rechazar una política o para defender una postura.

A esta clasificación se propone agregar un tipo más, cercano al instrumental, relativo a la utilización de las evaluaciones, no por la propia política o programa, sino por los órganos encargados de asignar recursos, muy importante en los países que ligan el desempeño al presupuesto (utilización presupuestal).

Existe un considerable debate sobre la medida en que la utilización se lleva o no a cabo, cuánto y cómo se puede aumentar. Una evaluación puede ser poco útil y en consecuencia no ser utilizada en ninguna forma, pero también puede incluir conclusiones pertinentes y de todas no utilizarse efectivamente para producir un cambio en la política o programa en cuestión. En consecuencia, aunque la calidad puede explicar en parte el nivel de utilización, otras muchas variables pueden incidir en el mismo.

Por otra parte, abundan los problemas metodológicos en el estudio de este tema. La mayoría de las investigaciones se han basado en el análisis de casos o en la aplicación de encuestas y entrevistas. Estas técnicas enfrentan dificultades para obtener documentación sobre la utilización, realizar su medición, demostrar la causalidad entre la evaluación y los cambios producidos, entre otras. Otro método sugerido es el análisis de contenido de una muestra representativa de evaluaciones.

Estados Unidos y Canadá durante el siglo XX

Debido a que los estudios existentes comparten los problemas antes enunciados, es fundamental la revisión de las variables que afectan la utilización de los resultados, especialmente del tipo instrumental. Durante el siglo XX existen algunos antecedentes relevantes, teóricos y empíricos, como las aportaciones de Weiss (1972 y 1988), Guba (1975), Patton (1978 y 1998), Leviton y Hughes (1981), Cousins y Leithgood (1986) y Mayne (1994) en el contexto de Estados Unidos y de Canadá. A continuación se realiza su revisión.

- Weiss (1972) define el concepto de utilización instrumental y en su trabajo de 1988 señala cuatro condiciones para que ésta se haga efectiva: 1) conclusiones de las evaluaciones que no estén controvertidas ni provoquen conflictos de interés al interior de las organizaciones, 2) recomendaciones que se encuentren dentro de las atribuciones del programa y no provoquen grandes rupturas en su funcionamiento, 3) ambiente del programa relativamente estable, y 4) momento por el que pasa el programa (si está en crisis y no se sabe cómo mejorarlo, esto puede favorecer la utilización de la evaluación).

- Patton (1978) escribió su conocido libro *Utilization-focused evaluation* poniendo énfasis en incrementar la utilidad de la evaluación, tanto en cantidad como en calidad, y considerando tanto los hallazgos como el proceso mismo dirigido a informar las decisiones y mejorar el desempeño de los programas. Para ello consideró necesario identificar con toda claridad a los usuarios principales de la evaluación, con objeto de focalizar el trabajo en ellos para que guiaran las decisiones durante todo el proceso, en lugar de que los evaluadores actuaran de forma independiente. Su idea era que los usuarios, al sentirse involucrados, comprenderían y sentirían como suya a la evaluación y usarían más sus resultados. De esta manera el evaluador prepararía el camino para que la evaluación fuera efectivamente usada.

Para lograr lo anterior, Patton empezó por sugerir cinco pasos: 1) identificar los principales usuarios potenciales, 2) comprometerlos con su enfoque de evaluación, 3) decidir conjuntamente sobre las opciones de evaluación posibles, 4) analizar e interpretar, también conjuntamente, los resultados y conclusiones, y 5) difundir los hallazgos.

Veinte años después, al preparar la tercera edición de su libro, Patton (1998) confesó que sus expectativas no se habían cumplido ya que en los últimos 10 años la utilización no había avanzado y los usuarios de sus propias evaluaciones manifestaban que lo relevante no eran los resultados logrados sino el proceso mismo, las vivencias de la gente y de las organizaciones como consecuencia de estar involucradas en un proceso de evaluación, que denominó “proceso de uso”, con su propio impacto más allá de los hallazgos concretos.

Trabajando en esta perspectiva, Patton fue ampliando los pasos incluidos como *check list* en su esquema de trabajo hasta alcanzar 17 en la cuarta edición de su libro, que van desde la planeación del trabajo de evaluación hasta la metaevaluación, incluyendo varios pasos especialmente dirigidos a incrementar las posibilidades de uso de la evaluación como: explicar las controversias e implicaciones metodológicas para negociar los métodos con los usuarios potenciales, aumentar la credibilidad de los hallazgos a obtener y consecuentemente su uso, simular las consecuencias de la utilización de los resultados para facilitar su implementación, metaevaluar el uso efectivo y el aprendizaje realizado, entre otros. Recientemente, Patton (2011) ha elaborado un pequeño libro con el esquema y los pasos esenciales de su método, presentándolo

como un sistema adaptativo complejo, incorporando elementos del pensamiento sistémico y conceptos de complejidad.

- En cuanto a Leviton y Hughes (1981), ellos ya señalan la importancia de los siguientes factores como explicativos del grado de utilización de los resultados de las evaluaciones:

- a) Relevancia.* Diversos autores (Averch, 1975; Williams y Wysong, 1975; Banta y Bauman, 1976; Cox, 1977, citados en Leviton y Hughes, 1981) han señalado que una mayor relevancia del contenido de la evaluación debería aumentar su utilización; sin embargo, no hay estudios que lo demuestren en forma generaliza, sólo algunos casos. Por ejemplo, cuando las evaluaciones responden a necesidades de algún grupo en particular son más utilizadas, pero si sólo miden la eficiencia de un programa en forma general son poco tomadas en cuenta por sus responsables y operadores, que prefieren la información generada en temas más específicos como asignación de recursos o procesos de implementación (Brickell, 1978, citado en Leviton y Hughes, 1981). En todo caso, siempre es necesario demostrar la suficiencia de la información recogida, analizada y valorada. También se ha hecho hincapié en la importancia de la presentación de resultados a tiempo que permita valorar la relevancia de la evidencia ofrecida para ser considerada en la toma de una decisión. En síntesis, de este punto se concluye la conveniencia de dirigirse a quien necesita la información, ofrecer a los funcionarios datos específicos bien fundamentados y entregar en tiempo los informes, para mejorar la utilización de resultados.
- b) Comunicación.* Dos estudios demostraron la importancia de una buena comunicación entre los evaluadores y los usuarios potenciales de su trabajo (Glaser y Taylor, 1973 y Windle y Bates, 1974; citados en Leviton y Hughes, 1977). Ésta no se lograría a través de la jerarquía burocrática, por ejemplo, si los niveles directivos sólo leen resúmenes elaborados por sus subordinados que pueden presentar errores o distorsiones, o si no transmiten la información generada por la evaluación a los niveles más bajos. Se recomienda, entonces, la comunicación directa con los usuarios potenciales de la evaluación (no la vía burocrática) y una difusión de los resultados también directa o por medio de redes.

- c) *Presentación y procesamiento de la información.* La forma en que se presentan los resultados de la evaluación a los usuarios afecta su comprensión y por lo tanto el grado de utilización. La confianza de los funcionarios aumenta si encuentran respuestas a sus preguntas específicas; en cambio, puede reducirse si el trabajo identifica efectos negativos no previstos, difíciles de asimilar. La utilización también se dificulta cuando los funcionarios rechazan la lectura de informes complejos porque prefieren una comunicación verbal y continua; lo que provoca que sólo perciban los resultados más vinculados a sus necesidades y difícilmente valoren críticamente la calidad de los trabajos. Leviton y Hughes (1977) anotan también que los administradores prefieren la información cualitativa a la cuantitativa que generalmente ofrecen los evaluadores y las observaciones a partir de vidas reales más que de experimentos en laboratorios. También que sus diferencias de estilo con los investigadores pueden afectar a la utilización. Este punto indica, en consecuencia, la necesidad de generar confianza en la relevancia del informe, presentarlo en forma clara y sin tecnicismos para los funcionarios, privilegiar la comunicación verbal y utilizar una combinación de métodos cualitativos y cuantitativos.
- d) *Credibilidad.* Esta variable depende de cuatro grupos de factores: sus coincidencias o diferencias con otras fuentes de información, las preconcepciones de los usuarios, el prestigio de los evaluadores y la calidad del trabajo. Si los resultados de la evaluación son sorprendentes a la luz de otras fuentes de información, de las expectativas o de la intuición y prejuicios del usuario, son menos susceptibles de ser utilizados. La credibilidad del evaluador puede reducirse por la sospecha de que haya sido cooptado o haya omitido alguna información (Guba, 1975). También se afirma que la calidad influye en el uso (Cook y Polard, 1977, citados en Leviton y Hughes, 1981), pero que no todos los autores están de acuerdo. Es probable que lo haga principalmente a partir del aumento de la confianza de que los resultados ofrecen una imagen precisa del programa.
- e) *Promoción del trabajo y participación del usuario.* Este apartado alude al amplio acuerdo en el sentido de que la promoción del trabajo es parte esencial del contexto político que explica la utilización de una evaluación, complementada con el compromiso del usuario con las evaluaciones, que implica la realización de una discusión seria de los resultados, no sólo una lectura del informe. El

compromiso asumido por el tomador de decisiones podría ser la variable crucial que contribuya a que las conclusiones se utilicen. Sin embargo, este compromiso resulta limitado por dos realidades burocráticas: el cambio de políticas de corto plazo y la rápida rotación entre los funcionarios de nivel superior en el gobierno, que puede provocar que los hallazgos de la evaluación lleguen demasiado tarde. Además, los funcionarios tienden a defender sus programas y, en consecuencia, se interesan sólo en las evaluaciones que muestran su éxito.

La falta de consenso en torno a las variables explicativas de la utilización de resultados presentadas por Leviton y Hughes (1981) muestra caminos heurísticos para continuar con la investigación, iniciada como se ha podido ver, hace alrededor de 40 años.

- Cousins y Leithgood (1986) contribuyeron con 65 estudios empíricos relativos a evaluaciones realizadas entre 1971 y 1985, mediante búsquedas computarizadas a partir de palabras-clave como “evaluation utilization”, “data use”, “decision making” o “knowledge utilization”, complementadas con la revisión de manuales y revistas relevantes. Organizaron los casos encontrados en dos categorías (utilización referida a procesos de implementación o a procesos de decisión), cada una con seis características. Las relativas a la primera categoría fueron: calidad de la evaluación, credibilidad, relevancia, calidad de la comunicación, hallazgos y tiempos. Las relativas a la decisión incluyeron: necesidades y disponibilidad de información, características de la decisión, clima político, características personales y compromiso con la evaluación. Identificaron a la calidad de la evaluación como la característica más importante, seguida de las características de la decisión, la receptividad a la evaluación, los hallazgos y la relevancia.
- Algunos años después, Mayne (1994) escribiría otro insoslayable trabajo sobre el tema en el contexto canadiense federal, que cuenta con unidades de evaluación interna en sus departamentos. Insiste en que, por lo general, no es posible confirmar definitivamente el impacto “real” de la evaluación en una organización o programa debido a los serios problemas de medición y atribución que se enfrentan. Sin embargo, sostiene que la utilización de la evaluación es mayor cuando se busca el equilibrio entre la demanda

de información de la evaluación y su oferta; pero también señala que las evaluaciones no tienen lugar en un entorno controlado, por lo que hay muchos otros factores que pueden afectar su aplicación, incluso facilitando su uso.

Gran parte de la literatura de utilización se basa en un modelo de evaluación externa, donde los evaluadores son contratados para llevar a cabo una determinada evaluación. Hay menos estudios en materia de evaluación interna, razón por la que el modelo canadiense le ofreció a Mayne la oportunidad para estudiar el grado en que las organizaciones y programas pueden aprender y el papel que la evaluación continua puede desempeñar en ese proceso.

El éxito de la utilización lograda es sintetizado para 1984-1991 mostrando que sólo entre 2 y 5% de las evaluaciones no tuvieron consecuencias, mientras que, en orden de importancia, el resto provocó: modificaciones al programa, una mejor comprensión del mismo, reformas, confirmaciones y terminaciones, acompañadas de reasignaciones presupuestales. Esto implica utilizaciones instrumentales, presupuestales y también conceptuales.

Para aprender del proceso anterior, que contradice la mayor parte de la literatura sobre utilización, se trató de lograr una mejor comprensión de los factores que contribuyen a que la utilización sea exitosa. Para ello se aplicó una encuesta a los 15 directores de evaluación más experimentados del gobierno federal sobre las expectativas de utilización, seguida de una discusión del tema. Las expectativas se concentraron en mejorar el desempeño del programa, pero no hubo pleno consenso al respecto. En la discusión quedó claro que las evaluaciones varían en función de múltiples factores; pero globalmente se consideró que el éxito dependía de que se satisficieran las necesidades departamentales. Los factores explicativos enunciados por los funcionarios fueron reunidos en tres grupos: factores organizacionales, proceso de evaluación y credibilidad de sus hallazgos. No se encontró un patrón respecto del peso que dichos funcionarios atribuían a cada grupo, incluso hubo quienes consideraron que el tercero era consecuencia de los dos anteriores.

Dentro del primer grupo los factores fueron: localización del equipo de evaluación en la estructura de la organización; nivel de aceptación general de la evaluación, factores político-burocráticos, personas específicas involucradas en el trabajo e importancia del programa evaluado. Salvo el último, todos los factores fueron considerados esenciales o, por lo menos, importantes.

En el segundo grupo se identificaron: conocimiento de evaluación e involucramiento del nivel gerencial superior, oportunidad de los resultados, participación de los beneficiarios, duración del trabajo, empleo de comités de asesores y consultores, entre otros. Nuevamente, todos menos el último fueron considerados muy importantes.

En cuanto a la credibilidad fueron mencionados principalmente las específicas conclusiones alcanzadas y el grado en que se fundamentan en la evidencia ofrecida, la confianza en el evaluador y el rigor metodológico (este último, aunque importante, no afectaría tanto la credibilidad). En menor grado se aludió a la comunicación organizacional, la importancia de los hallazgos para la organización y la participación de expertos externos.

Proliferación mundial de estudios en el siglo XXI

En años más recientes aparecen numerosos trabajos con la misma preocupación y en contextos muy diversos como Reino Unido, Estados Unidos, Canadá, Finlandia, Dinamarca, Suiza, Francia, España, Australia, Nueva Zelanda, Japón, entre otros.¹ La bibliografía disponible en francés y español, salvo mínimas excepciones como la evaluación del aprendizaje en la educación o las discusiones desarrolladas en el seno de la SFE, sólo analiza los procesos de evaluación hasta la emisión y difusión del informe, en consecuencia, no se cuestiona sobre su utilización posterior. A continuación mencionamos algunas contribuciones de los países señalados. En 2002 se hace referencia a cuatro trabajos, dos relativos a aumentar la utilidad potencial, otro que busca ampliar el concepto más allá de su instrumentalidad y uno más que analiza los factores empíricos que facilitan la utilización:

¹ Boaz y Hayden (2002), en Reino Unido, Valovirta (2002) en Finlandia, Leviton (2003) en Estados Unidos, Yoshizawa (2009) en Japón, Balthasar (2009) en Suiza, Johnson y otros (2009) en Estados Unidos, Jacobson (2011) en Australia, Saunders (2012) en Reino Unido, Patton (2012) en Estados Unidos, Contandriopoulos (2012) en Canadá, Appleton-Dyer y otros (2012) en Nueva Zelanda, Hojlund (2014) en Dinamarca.

- Annette Boaz y Carol Hayden (2002), en el Reino Unido, constatan el relativo abandono de modelos racionales basados en evidencias que se han ido sustituyendo por una perspectiva “realista”, que requiere que el evaluador promueva su contribución potencial, lo que implica que se vuelva proactivo y desarrolle mayores capacidades para comunicar sus resultados, con el fin de aumentar la probabilidad de que sean utilizados.
- Lille Valovirta (2002), en Finlandia, insiste en la perspectiva argumentativa del evaluador puesto que la información generada por la evaluación no puede ser considerada un conocimiento indiscutible, por lo que propone un esquema de análisis que resalta dos dimensiones (grado de presión hacia el cambio y relación entre niveles de conflicto y consenso) que se ejemplifican en agencias gubernamentales de su país.
- Feinstein (2002), investigador del BM, parte de insistir en el muy limitado uso de las evaluaciones para analizar los factores económicos y políticos que afectan su utilización: oportunidad, credibilidad, presentación del informe y canales de distribución.
- Forss *et al.* (2002), en los países escandinavos (Suecia, Dinamarca), hablan de la utilización no sólo instrumental y *a posteriori*, sino de las lecciones aprendidas durante el proceso (*process use* de Patton), cuyo reconocimiento explícito puede aumentar la relación costo/efectividad de la utilización.
- Por otro lado, y como parte de las excepciones en el contexto francés, Perret (2008) se pregunta si es posible evaluar el impacto de la evaluación ya que enfrenta los mismos problemas técnicos que la evaluación de programas, especialmente la atribución del cambio observado. Afirma que a nivel internacional todos los estudios disponibles señalan que el impacto más importante es lento, indirecto y difuso (no instrumental e inmediato), y que se produce principalmente mediante cambios en las maneras de pensar el programa, en la clarificación de objetivos, en la reducción de incertidumbres, en el reforzamiento de compromisos individuales y en el conocimiento que llega indirectamente a los decisores a través de las discusiones en la prensa o entre universitarios. Sostiene que, al menos, podemos disponer de algunos datos interesantes que muestran que queda mucho camino por andar. Aunque la evaluación ha aportado notablemente a la reforma gerencial en casos como Reino Unido, los países que más la utilizan de manera instrumental la relacionan con la preparación del presupuesto (Estados Unidos,

Canadá, Reino Unido, Países Bajos y Suecia) y cuentan con una cultura política que otorga un peso relevante al interés de los contribuyentes. En dichos casos existen menciones específicas a las evaluaciones para justificar cambios presupuestales (65% de ellos en el caso de Australia en 1994-1995). La situación antes descrita no existe en Francia, donde el carácter político de la discusión presupuestal hace que la influencia instrumental de la evaluación sea limitada, lo que, en opinión de Perret, constituye la principal causa de su débil utilización. Pero también en sentido amplio su impacto es limitado ya que las evaluaciones interministeriales, que han sido ampliamente publicadas, no han tenido eco en los ciudadanos a pesar de tratar temas cruciales como el empleo.

Por último, Perret realiza una precisión muy importante: incluso cuando la evaluación es usada, la finalidad puede ser desvirtuada para lograr objetivos políticos como reforzar retóricamente la credibilidad, la legitimidad y la autoridad del discurso de los poderes públicos. La propia OCDE (1997, citada por Perret) ha señalado que la evaluación puede ser realizada con el objetivo de legitimar decisiones políticas ya tomadas, puede eliminar evidencias incómodas o políticamente sensibles y utilizar la información de manera selectiva. Perret concluye que una adecuada institucionalización de la evaluación, como se ha intentado en la Unión Europea, podría contribuir a reducir estas manipulaciones.

- Johnson *et al.* (2009) ofrece una importante revisión de 41 estudios empíricos de uso efectivo de evaluaciones desarrolladas entre 1986 y 2005, la gran mayoría de éstos en Estados Unidos, siguiendo el esquema de dos categorías y 20 características propuesto por Cousins y Leithgood (1986) y añadiendo una categoría (participación de interesados) y una característica (competencia del evaluador) adicionales. También replicaron su metodología, incluyendo la revisión de 600 informes, capítulos y artículos y 48 tesis en lengua inglesa, entre los que a partir de diferentes criterios aplicados terminaron por seleccionar 41 textos completos a analizar, seis de ellos publicados fuera de Estados Unidos. Por último, midieron la variable dependiente “nivel de uso” y las diversas variables explicativas del mismo. Concluyeron que la participación de los diversos actores involucrados aumenta considerablemente el uso de la evaluación y que el compromiso, interacción y comunicación entre los evaluadores y sus clientes es crítico para que esta utilización sea significativa

en el largo plazo. El uso instrumental fue más frecuente que el conceptual, 38 de los 41 casos se enfocaron a la utilización de los hallazgos y sólo tres al proceso de uso.

- Contandriopoulos y Brousselle (2012) en Canadá, quienes sostienen que el rol de los evaluadores y el proceso desarrollado permiten predecir su uso, resaltando que el contexto de la evaluación condiciona tanto la apropiación del modelo aplicado como la utilización de los resultados obtenidos.
- Appleton-Dyer *et al.* (2012) en Nueva Zelanda, parten del reconocimiento de que la utilización de la evaluación es un proceso complejo, aún no bien explicado, por lo que se limitan a señalar un conjunto de factores que los evaluadores pueden tomar en cuenta en su práctica para mejorar la influencia de su trabajo, como los atributos de la evaluación misma, el funcionamiento y características de los actores involucrados, su conducta, las características individuales y los factores contextuales.
- Bonbright (2012), en Estados Unidos, es la cuarta de una serie de guías preparadas por InterAction y está focalizada en la utilización de los resultados de las evaluaciones de impacto, que es calificada como una práctica a lo sumo emergente. Plantea que su utilización aumenta si se planea conjuntamente con la evaluación o, por lo menos, desde sus primeras etapas, si las organizaciones responsables cuentan con los sistemas operativos necesarios y, especialmente, con una cultura favorable a la evaluación. Para lograrlo sugiere crear un sistema de incentivos para la adopción de los hallazgos.
- Hojlund (2014) en Dinamarca, también parte de la paradoja que la evaluación se haga para mejorar los resultados pero rara vez se utilice. Se cuestiona sobre las razones que la provocan y sugiere utilizar la teoría organizacional del neoinstitucionalismo (función simbólica, legitimidad) para encontrar la explicación en función de presiones externas y propensión interna a la evaluación, en lugar de seguir buscándola en la evaluación misma (su implementación, resultados, etcétera), que además no considera plenamente institucionalizada. Para ello clasifica la utilización en cuatro posiciones ideales y extremas que caracteriza en detalle: adopción coercitiva (baja propensión interna y alta presión externa), mimética (alta en los dos factores), normativa (baja en los dos) y voluntaria (alta interna y baja externa).
- Laubli y Mayne (eds.) (2014) ofrecen una nueva compilación de trabajos entre los que destacan: el propio de Mayne, identificando la importancia

crítica de los factores organizacionales en la utilización de los resultados de la evaluación; el de Laubli proponiendo una adaptación creativa de la evaluación en el contexto de la Unidad de Evaluación de la Oficina Federal Suiza de Salud Pública; el de Laat y Williams que sintetiza los hallazgos de dos estudios empíricos, en 2002 y 2005, sobre el uso de las evaluaciones en la Comisión Europea, una de las organizaciones mundiales que más evaluaciones contrata; y el de Neubecker *et al.*, que presenta las técnicas innovadoras usadas en la Organización Internacional del Trabajo (OIT) para promover la utilización de los resultados de evaluación.

Entre los últimos citados, Laat y Williams identifican cinco tipos de utilización de resultados y presentan la frecuencia de casos encontrados: 8 en implementación, 7 en diseño, 4 en asignación de recursos, 4 en rendición de cuentas y 2 en establecimiento de prioridades y elección entre opciones de políticas. La mayoría de los anteriores (salvo rendición de cuentas) se vinculan con una utilización instrumental; le sigue el desarrollo de conocimientos de base y luego el aprendizaje y la disposición de elementos para el debate de políticas. Los autores también analizan los factores que afectan dicha utilización: planeación y tiempos del proyecto de evaluación, apoyo de la alta dirección, calidad de la evaluación, participación de los usuarios potenciales y beneficiarios, difusión de sus resultados, y un sistemático y transparente monitoreo de la implementación de las recomendaciones.

España y América Latina

- Como parte de las pocas excepciones que tratan el tema en lengua española, en el contexto de la CEPAL, Wiesner (2000) recupera la idea de que las evaluaciones en América Latina, son tomadas en cuenta cuando se las vincula con los procesos presupuestarios. En ese sentido Neirotti (s/f) señala que esa relación está establecida por ley en el caso de Chile y en la propia Constitución en el de Brasil; sin embargo, sus niveles de obligatoriedad no están claramente señalados, resultando más articulados los casos de Costa Rica y Colombia, mientras la relación mencionada está completamente ausente en Argentina.

- Zapico (2005) alude al tema de la relación entre evaluación y presupuesto por resultados, agregando referencias al contexto español y latinoamericano (Chile). Comienza por reconocer los lentos avances logrados en la asignación de recursos en función del desempeño alcanzado, para pasar luego a analizar los principales obstáculos propios de su complejidad creciente, al que los modelos tradicionales de evaluación (aplicación mecánica de indicadores operativos y cuantitativos) y presupuesto no se ajustan correctamente: pluralidad de intereses a menudo conflictivos, incertidumbre respecto de las metas y objetivos, interdependencia entre múltiples organizaciones y niveles de gobierno implicados en programas conexos, inestabilidad y cambio económico y social, entre los principales. Por otro lado, señala como requisitos para su avance el contar con: un liderazgo continuado, voluntad política y apoyo directivo, una clara distribución de responsabilidades y construcción de confianza entre los actores involucrados, una amplia capacidad de aprendizaje por parte de la administración pública, así como la introducción gradual de la evaluación y el presupuesto por resultados, basada en un diagnóstico específico y acompañada de incentivos para el uso de la información generada.
- También desde el contexto español, Bustelo (2013) se refiere a la credibilidad como requisito para el uso de las evaluaciones, resaltando los principales problemas que la afectan: especialización e independencia de los evaluadores, grado de transparencia del proceso de evaluación, calidad de los productos, así como su contexto cultural, incluyendo los valores sobre los que se fundamenta la evaluación, la credibilidad de las evidencias, las instituciones que apoyan los sistemas y estructuras de evaluación, las personas que contribuyen a realizarla y la comunicación y presentación del informe.

La autora afirma que, aunque su proceso se desarrolle en un contexto político, las evaluaciones, además de ofrecer evidencia sistemática y rigurosa, necesitan ser percibidas como imparciales. Si bien a menudo el tiempo y los recursos disponibles limitan su uso, considera que un enfoque participativo promueve la apropiación de resultados por sus interesados. También impacta la credibilidad la metodología, los datos recogidos y analizados, el papel que desempeñan las instituciones en la organización de los procesos de evaluación, su localización en el Poder Ejecutivo o Legislativo y la capacitación y experiencia de los evaluadores y de los funcionarios que solicitan sus trabajos.

Bustelo parece referirse más a la credibilidad social que a la de los funcionarios que toman decisiones.

Finalmente, en el contexto mexicano han transcurrido aproximadamente 15 años de experiencias gubernamentales que han sido acompañadas por una reducida discusión académica sobre cómo mejorar la calidad de las evaluaciones, pero es hasta hace muy poco tiempo que el tema de la utilización de sus resultados ha aparecido como el talón de Aquiles de nuestros sistemas. En ese sentido se recuperan a continuación los pocos trabajos disponibles.

- Cejudo y Maldonado (2011) compilaron un libro que presenta un conjunto de trabajos de revisión de este asunto en programas sociales, educativos, de salud y ambientales, en el marco del Premio a los Programas Federales Comprometidos con el Proceso de Evaluación, instituido por el CIDE, la SFP y la asociación civil México Evalúa para reconocer a los programas que han implementado acciones recomendadas dentro del proceso de evaluación (sin valorar la calidad de los programas y evaluaciones mismas) y para incentivar una mayor utilización.² Cejudo reconoce un importante avance del sistema de evaluación pero también una larga serie de obstáculos para que sus resultados se utilicen. Cada recomendación tiene la posibilidad de generar una acción de mejora, pero que ésta se concrete depende de diversos factores: la capacidad del evaluador para identificar las oportunidades de mejora, el conocimiento que tenga del programa para proponer recomendaciones relevantes, el nivel técnico, interés y capacidad administrativa de los decisores para entender los hallazgos, diseñar modificaciones y llevarlas a cabo, entre otros. El instrumento ASM puede ser utilizado para internalizar la lógica de las evaluaciones en la administración pública y el comportamiento de los funcionarios, pero también puede suceder que éstos sólo se apropien de los procedimientos y no de la sustancia, que es lo más importante. El autor señala que buena parte de los obstáculos para la utilización de los resultados proviene de algunas características del sistema mismo, como la estandarización de las evaluaciones, sus tiempos y formatos (no

² Fueron seleccionados 10 programas de los 14 que se presentaron a la convocatoria.

necesariamente deseables), o la preocupación de los funcionarios por cumplir en “tiempo y forma” más que por generar valor público y, en consecuencia, su interés en que las evaluaciones arrojen resultados positivos más que en aprender y mejorar a partir de sus críticas.

Las características anteriores afectan la calidad de la evaluación que depende de los procedimientos de contratación, la evaluabilidad del programa, la calidad y confiabilidad de la información y la solidez técnica y metodológica del evaluador. Además, las recomendaciones han de ser pertinentes, factibles, claras y precisas para que sean utilizables. Pero aun cumpliendo la evaluación con estos requisitos, falta que sus resultados sean aceptados e instrumentados en los propios programas, donde aparece otro de los obstáculos ya que persiste la idea de que la evaluación es similar al control y de poca utilidad práctica. En este contexto de dificultades resaltan las experiencias aisladas tendientes al desarrollo de la cultura de la evaluación que se relatan en el libro.

- Montiel (en Cejudo y Maldonado, 2011), cuyo análisis se enfoca fundamentalmente a las ECR de 106 programas sociales federales en 2007-2008 y a la realización de entrevistas con quienes presentaron un cumplimiento pleno (no representativas), también reconoce el carácter informativo y simbólico de las evaluaciones junto con su falta de utilización práctica y nuevamente se pregunta sobre los factores que la explican. Entre éstos incluye la falta de incentivos y un diseño institucional débil, caracterizado por: atribuciones compartidas entre Coneval, SHCP y SFP; presupuesto reducido del primero en comparación con las otras entidades reguladoras, poco personal, escasas atribuciones para sancionar los incumplimientos, entre otras.

Por otro lado, entre los factores mencionados por los entrevistados como explicativos de su nivel de cumplimiento se encuentran: 1) el liderazgo, que constituyó un factor facilitador en el caso de CDI, 2) la calidad de la evaluación, que inhibió su utilización en Semarnat, 3) la claridad de las reglas, que obstaculizó el cumplimiento de Conagua, 4) las relaciones entre instancias administrativas al interior de la dependencia, cuya descoordinación también afectó el incumplimiento de Semarnat, y 5) el tiempo que, en general, se considera insuficiente para lograr un pleno cumplimiento.

Con base en lo anterior, Montiel concluye que el sistema de evaluación mexicano muestra indicios de funcionalidad pero que no retroalimenta el

ciclo de políticas públicas para mejorar los programas y lograr mejor calidad de vida de la población.

- También resaltando la falta de vínculo entre la etapa terminal (evaluación) de un ciclo de política pública y la retroalimentación del siguiente (reformulación), Cardozo (2013) analiza el nivel de utilización de las evaluaciones coordinadas en México tanto por Coneval a nivel federal, como por Evalúa DF en el caso de la capital del país, y en la asignación de presupuesto por parte de la Cámara de Diputados, así mismo en relación con su uso instrumental para la modificación de políticas y programas sociales. Además, considerando que uno de los factores explicativos del nivel de uso de las evaluaciones se vincula con su calidad, se revisan los procesos de metaevaluación iniciados por Evalúa DF y otros que persiguen un fin similar en el contexto del Consejo Nacional de Evaluación de la Política de Desarrollo Social.

Encuentra que, si bien los niveles de cumplimiento son considerables (alrededor del 73-74%) para el tiempo que ambos Consejos llevan trabajando, éstos son aún insuficientes y se concentran en ajustes a la operación de los programas, afectando en menor grado su reformulación sustantiva. Esta similitud de resultados se presenta a pesar de las importantes diferencias jurídicas, organizacionales, de enfoques y de métodos presentes en los dos casos, dentro de las que resalta el carácter inductivo de los ASM del Coneval frente a las recomendaciones vinculatorias de Evalúa DF. Por lo que concluye que no existe evidencia suficiente de que las evaluaciones mexicanas estén contribuyendo a un mejor desempeño de los programas sociales, debido a problemas contextuales, organizativos y metodológicos.

Síntesis

A continuación se sintetizan en un cuadro los múltiples factores explicativos de esta paradoja (las evaluaciones se realizan para mejorar las políticas y los programas, pero éstos las recuperan muy poco) que han sido señalados por los autores revisados. Obviamente, para concentrar su gran diversidad ha sido necesario agruparlos en función de sus contenidos y para facilitar su comprensión se han organizado en torno al proceso teórico seguido por las evaluaciones (Cuadro 33).

En el cuadro mencionado se reúnen autores de contextos nacionales diferentes y variadas formas de organización de la evaluación. En torno de los factores con mayor poder explicativo del nivel de utilización reina una fuerte diversidad y falta de consenso pero, en términos generales, nos permite visualizar la concentración de referencias teóricas y empíricas a los temas de cultura y aceptación de la evaluación como prerrequisito fundamental para la utilización efectiva de sus resultados, así como a la importancia del involucramiento de sus usuarios a lo largo del proceso para mejorar su apropiación de los hallazgos.

En segundo lugar se ubican las menciones vinculadas con la credibilidad de la evaluación, que si bien se relaciona con su calidad, parece depender más del prestigio de los evaluadores. Esta relativa importancia de la calidad de la evaluación se refuerza con las pocas referencias encontradas a la rigurosidad metodológica del trabajo realizado.

En una posición cercana a la anterior se ubican el contexto político y normativo, el diseño de interdependencias gubernamentales y la estabilidad del programa que pueden favorecer u obstaculizar la utilización de las recomendaciones. De igual forma aparecen en este sitio los aspectos organizativos: necesidad percibida de información sobre el programa, nivel jerárquico solicitante, apoyo de autoridades superiores y planeación del proceso de evaluación, incluyendo desde el inicio la difusión y utilización efectiva de los resultados.

Si bien muchos de los trabajos revisados mencionan la vinculación entre una alta utilización de resultados y los procesos de asignación de recursos presupuestales, el cuadro muestra muy pocas menciones expresas al diseño de un sistema de incentivos para la adopción de las recomendaciones, lo que podría ser un camino alternativo posible para México, mientras avanza el lento proceso de cambio cultural en la materia.

CUADRO 33
Principales factores explicativos de la utilización de las evaluaciones
(autores y año de publicación)

Autores y año de su publicación	a	b	c	d	e	f	g	h
Patton (1978; 1998)			X		X			X
Levington y Hughes (1981)		X	X		X		X	X
Cousins y Leithgood (1986)		X	X			X	X	
Weiss (1988)	X		X			X		
Mayne (1994)	X	X	X			X	X	
Boaz y Hayden (2002)			X					
Valovirta (2002)			X			X		
Feinstein (2002)					X		X	X
Zapico (2005)	X	X	X	X				
Perret (2008)	X							X
Johnson <i>et al.</i> (2009)			X					
Cejudo y Maldonado (2011)	X	X	X	X		X	X	
Contandriopoulos y Brousselle (2012)	X				X		X	
Appleton Dyer <i>et al.</i> (2012)	X						X	
Bonbright (2012)		X	X	X		X		
Bustelo (2013)		X	X		X		X	X
Cardozo (2013)	X	X	X			X	X	X
Hojlund (2014)	X		X					
Laubli y Mayne (eds.) (2014)		X	X				X	X

- a) Clima político, normatividad, interdependencias gubernamentales, estabilidad del programa
- b) Necesidad de información, nivel jerárquico del solicitante y planeación de la evaluación y su utilización
- c) Cultura de evaluación y liderazgo, comunicación y participación de los usuarios potenciales
- d) Sistemas de incentivos para la adopción de recomendaciones
- e) Fuentes de información, procesamiento y presentación de la información
- f) Conclusiones fundamentadas, discutidas, no controvertidas, y recomendaciones factibles
- g) Credibilidad, prestigio de evaluadores, calidad, relevancia y oportunidad del informe
- h) Canales de distribución y difusión del informe y monitoreo de su uso

Fuente: elaboración propia con la información revisada en el capítulo.

En síntesis, la revisión teórica efectuada permite concluir que la tendencia mundial muestra poca utilización de los resultados de la evaluación con algunas excepciones consolidadas en Estados Unidos, Canadá y Reino Unido y otras incipientes en países latinoamericanos como Costa Rica o Chile.³ Los autores revisados señalan un amplio y disperso conjunto de posibles factores explicativos de este fenómeno que incluye las condiciones del contexto, la cultura de evaluación, la credibilidad del trabajo vinculada especialmente al prestigio de los evaluadores, la participación de los usuarios potenciales de los resultados (funcionarios del programa) y los aspectos organizativos del procesos de evaluación. En cambio, la calidad del informe y la rigurosidad metodológica del trabajo sólo aparecen mencionadas en pocos casos.

³ Otros casos importantes mencionados por Perret son Países Bajos y Suecia, pero su experiencia no fue analizada en este libro.

CAPÍTULO XXII

Análisis de experiencias

Myriam Cardozo Brum

Países exitosos

Actualmente EvalPartners está realizando un concurso mundial para publicar en 2015, Año Internacional de la Evaluación, experiencias en este tema que, además de ser de buena calidad o que hayan contribuido a mejorar sus programas (utilización instrumental), hayan logrado un impacto significativo para mejorar las condiciones de vida de la población. Por dicha razón la convocatoria se dirige más a recoger la perspectiva de los usuarios que la de los evaluadores. Además, difiere de los informes de evaluación o de los artículos publicados en revistas académicas, puesto que busca “contar una historia”, cuyos atractivos sean tanto intelectuales como emotivos.

Recientemente se han dado a conocer los programas ganadores, que corresponden a nueve países: Canadá, El Congo, Indonesia, Kenia, Holanda, Nueva Guinea, Nepal y Sri Lanka y México (con el Programa Oportunidades). Si bien la convocatoria incluyó a todos los países agrupados en seis regiones (Europa, América Latina y el Caribe, Norteamérica, África, Asia y Australasia), evidentemente la selección casi no incluye la experiencia de países avanzados en materia de evaluación, sino que se concentra en aquellos de más reciente inmersión en dicho proceso.

Por otra parte, con base en la información presentada en la primera parte del libro, se mencionan a continuación los países que mostraron mayor nivel de utilización, fundamentalmente instrumental, presupuestal y persuasiva.

- Estados Unidos de América. La evaluación se utiliza intensamente con fines persuasivos (o disuasivos) para defender (o atacar) políticas, tanto por el Poder Ejecutivo Federal como por los gobernadores de los estados y el Congreso, en un contexto de fuerte debate político. Al interior del primero, la preocupación por incrementar su utilización fue el fundamento de la GPRAMA de 2010, y en el último se concentra en el uso de los resultados evaluativos de la Government Accountability Office.
- Reino Unido. Ha utilizado intensamente la evaluación con objetivos presupuestales desde fines de los años sesenta, tanto a nivel de los Ministerios de Economía y Finanzas, como en la formulación de políticas que requieren evidencia fiable. El Parlamento utiliza la información generada por la NAO sobre el gasto público.
- Canadá. Cuenta con un fuerte sistema de evaluación intragubernamental que utiliza en forma instrumental la información generada por sus propias evaluaciones sin necesidad de recurrir a ningún mecanismo coercitivo, sino basado en un prolongado desarrollo de una cultura institucionalizada entre los funcionarios.
- Costa Rica. Aunque muestra un reducido número de evaluaciones anuales, su nivel de utilización instrumental de las recomendaciones es elevado en la elaboración del PND y en el mejoramiento de los servicios públicos. Al igual que Canadá, pone énfasis en la autoevaluación de los funcionarios pero, a diferencia de aquél, los somete a un proceso de negociación para la firma de los Compromisos de Resultados que acompaña de un sistema de incentivos a la utilización de los resultados de las evaluaciones.
- Chile. Inició la utilización analizada en torno al presupuesto y la eficiencia del gasto pero recientemente ha creado incentivos individuales y colectivos (monetarios y reconocimientos simbólicos) para que la evaluación se utilice de manera efectiva como herramienta para mejorar la gestión pública, sin emplear mecanismos obligatorios y punitivos. A partir de 2012, si la institución alcanzó un grado de cumplimiento igual o superior al 90% de los objetivos anuales comprometidos recibe un incentivo colectivo que alcanza al 7.6% de las remuneraciones por unidad, y si logró más del 75% pero menos del 90%, un 3.8% (Dipres, 2012). También existen mecanismos más coercitivos, similares a los de Costa Rica y a los ASM mexicanos, donde

las instituciones firman Compromisos Institucionales para cumplir con las recomendaciones de evaluación en un plazo determinado.

México

En este caso se identifican tres tipos de utilización de los resultados de las evaluaciones: *a*) el instrumental, que corresponde a los cambios retomados por el propio programa, *b*) el presupuestal, en que la evaluación aporta información e incluso elementos de persuasión para la asignación de recursos, y *c*) el conceptual, que permite mejorar los conocimientos manejados por evaluadores, funcionarios, consejeros, legisladores y ciudadanía en general. Su presentación se organiza en función de los niveles de gobierno: federal y del Distrito Federal.

A nivel federal

En este sentido, la mayor utilización es de tipo instrumental y en menor grado presupuestal, mientras la conceptual resulta de muy difícil medición, como se presenta a continuación.

- Utilización instrumental en el conjunto de programas

Se comienza con la revisión de la utilización instrumental con la intención de analizar si las recomendaciones son tomadas en cuenta por los responsables de políticas y programas para mejorar su diseño y planear mejor su implementación, con especial interés en el caso más relevante de los programas sociales (Oportunidades) y de algunas experiencias consideradas especialmente exitosas.

Al iniciarse los procesos de evaluación en el nuevo siglo, la Sedesol puso en marcha un Sistema de Seguimiento a Evaluaciones Externas (Sisee) que controlaba el cumplimiento de los compromisos asumidos por los programas mediante convenios pero, con la excepción de los programas de desnutrición, éste dejó lentamente de cumplir su cometido al transformarse en un proceso formal de entrega de informes de avances porcentuales de cumplimiento,

sin identificar las acciones sustantivas realizadas para resolver los problemas identificados (Cardozo, 2009).

A partir de 2006 en que se creó Coneval, puesto que sus recomendaciones no son vinculatorias, este Consejo diseñó conjuntamente con la SHCP y la SFP un sistema de control de gestión para hacer más efectivo el cumplimiento de los convenios. Actualmente, después de presentado el informe final de evaluación, la Unidad Responsable de la Operación del Programa emite un oficio dirigido al Coneval, la SHCP, la SFP estableciendo su Posición Institucional ante la evaluación, al que ya se hizo referencia en la parte II de este libro. Recibido éste, las tres instancias mencionadas seleccionan, con base en criterios de claridad, relevancia, justificación y factibilidad, los denominados ASM a acordar con sus operadores para mejorar la gestión del programa. Posteriormente se establece un Convenio de Compromisos de Mejoramiento de la Gestión que el programa debe instrumentar.

Los ASM se definen y se publican en la página oficial del Coneval desde la generación de evaluaciones de 2008-2009 y su Mecanismo de Seguimiento se aplica anualmente por la SFP. Actualmente son clasificados en función de su nivel de prioridad en:

- Aspectos específicos, cuya solución corresponde únicamente a las unidades responsables.
- Aspectos institucionales, que requieren la intervención de dos o más áreas de la dependencia o entidad.
- Aspectos interinstitucionales, que demandan la participación de más de una entidad o dependencia, y
- Aspectos intergubernamentales, que necesitan de la intervención adicional de otros ámbitos de gobierno (estatal y/o municipal).

En el caso de los aspectos interinstitucionales e intergubernamentales la LGDS otorga atribuciones para la toma de decisiones a la Comisión Nacional de Desarrollo Social (CNDS) y a la Comisión Intersecretarial de Desarrollo Social (CIDS).

Puesto que existe muy poca información del proceso de transformación de las recomendaciones en ASM, tanto en cantidad como en contenido, resulta

relevante analizar los datos presentados por el Informe de Seguimiento de ASM de 2013-2014, incluidos en el Cuadro 34.

El cuadro mencionado muestra que aproximadamente 45% de las recomendaciones se transforman en ASM, aunque su distribución por instituciones presenta casos extremos como el de la Sedesol con sólo 28% y, en sentido inverso, Inmujeres (incluido en otros), que a pesar de que no recibió recomendaciones presenta siete ASM. Lamentablemente no hay explicaciones de estos resultados ni explicitación de los procesos y criterios que conducen a ellos.

CUADRO 34
Número de recomendaciones y de ASM en la generación de evaluaciones 2012-2013

Dependencias y entidades	Número de recomendaciones	Número de ASM	Porcentaje de ASM respecto de las recomendaciones
Sedesol	206	58	28
Semarnat	103	39	38
SE	97	31	32
SEP	45	38	84
Salud	39	18	46
Sagarpa	29	17	59
CDI	10	10	100
Otros	41	49	120
Total	570	260	45

Fuente: elaboración propia a partir del Informe de Seguimiento de los ASM 2013-2014 (Coneval, 2014b).

A continuación se presenta un breve análisis de la evolución de los ASM y de su cumplimiento en el periodo de 2008 a 2013. El Cuadro 35 muestra un número fluctuante de instancias participantes y programas evaluados: 13 o 14 dependencias y entidades, alcanzado un máximo de 16 en 2012-2013 y una media de 112 programas anuales, cuya variabilidad depende mucho del tipo de evaluaciones realizadas, ya que por ejemplo las ECR son bastante más complejas que las evaluaciones específicas de desempeño.

En cuanto a la información contenida en el Cuadro 36, a pesar de que el número de programas se mantuvo relativamente estable entre 2008 y 2010

(119 y 108 programas respectivamente), se produjo una drástica disminución del número de ASM (de 930 a 413), con un leve aumento del porcentaje de los específicos y la correspondiente reducción de los interinstitucionales e intergubernamentales, así como de los de alta prioridad (Montiel, 2011). En años posteriores el número de ASM por programas volvió a aumentar en 2011-2012 (5.7) pero sin alcanzar el nivel de 2007-2008 (7.8). En 2013-2014 habría un aumento sustantivo respecto del año anterior al pasar de 260 a 375.¹

CUADRO 35
Número de programas evaluados, de instancias responsables y de ASM 2008-2013

	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	Total 2008-2013
Núm. de dependencias y entidades	11	15	14	13	13	16	82
Núm. de programas evaluados**	119*	150	108	130	113	80	672
Núm. de ASM fomulados	930	871	413	471	641	260	3 586

* Nota: el informe de 2008 menciona 91 programas evaluados en su página 4 y 119 en su página 15.

** Nuevamente este número de programas evaluados está muy lejos de coincidir con los programados en el PAE y los efectivamente realizados que fueron presentados en el Cuadro 11.

Fuente: elaboración propia con base en los Informes de Seguimiento de los ASM del Coneval, correspondientes al año siguiente al de la adopción de los ASM aquí reportados.

¹ Éste se calculó por diferencia ya que la respuesta del Coneval al cuestionario elaborado señala un total de 3 031 ASM en 2008-2014.

CUADRO 36
Distribución de ASM por tipos 2008-2013

	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013	Total 2008-2013	%
ASM Específicos	651	723	370	402	497	207	2 850	79.5
ASM Institucionales	204	105	17	61	129	51	567	15.8
ASM Interinstitucionales	28	26	16	6	11	1	88	2.5
ASM Intergubernamentales	47	17	10	2	4	1	81	2.2
ASM totales	930	871	413	471	641	260	3 586	100

Fuente: elaboración propia con base en los Informes de Seguimiento de los ASM del Coneval, correspondientes al año siguiente al de la adopción de los aspectos susceptibles de mejora.

Los datos complementarios más relevantes (Cuadro 37) confirman una vez más que el importante incremento en el número de evaluaciones realizadas desde la creación del Coneval se ha hecho sobre la base de una gran simplificación (EED), en detrimento de las evaluaciones de impacto, casi inexistentes, y de la esporádica concreción de ECR que serían las de mayor cobertura temática, aunque sin incluir trabajo de campo. Otro tema descuidado es el de la percepción de los beneficiarios que sólo se reporta en la ECR de 2007-2008. Mientras el número de este tipo de evaluaciones fue similar en sus dos aplicaciones (119 en 2007-2008 y 113 en 2011-2012), el de las EED ha tendido claramente a su reducción (de 150 en 2008-2009 a 80 en 2012-2013).

En relación con la concentración de ASM por instituciones, la SEP aparece en todos los años entre los primeros tres lugares, concentrando entre 33% en 2007-2008 y 14% en 2009-2010. Le siguen la Sedesol, la Semarnat y la CDI, lo que en conjunto se explica porque son las cuatro instancias con mayor cantidad de programas evaluados anualmente (aunque esta información no siempre aparece desglosada). Si tomamos en cuenta el informe que da seguimiento a los 113 programas evaluados en 2011-2012, la SEP contribuyó con 32% de los programas y 29% de los ASM; la Sedesol alcanzó 13% en ambas mediciones,

CUADRO 37
Datos complementarios sobre las evaluaciones realizadas y los ASM emitidos

Datos complementarios	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013
Tipo principal de evaluación	ECR	EED	EED	EED	ECR	EED
Porcentaje de concentración institucional de los ASM	SEP 33 Sedesol 24 SHCP 16	SE 27 SEP 25 Sedesol 16	Semarnat 28 SEP 14 CDI 12	SEP 29 CDI 18 Semarnat 14	SEP 29 CDI 21	Sedesol 22 Semarnat 15 SEP 15
Porcentaje de ASM de alta prioridad	66	58	55	58	38	41

Fuente: elaboración propia con base en los Informes de Seguimiento de los ASM del Coneval, correspondientes al año siguiente al de la adopción de los aspectos susceptibles de mejora.

mientras que la CDI fue responsable de 12% de los programas pero su porcentaje de ASM fue casi el doble (21%).

Montiel (2011) consideró que los reportes correspondientes a los tres primeros años analizados demostraban un mayor esfuerzo por parte del Coneval en la presentación de información. Subrayaba que las reglas sobre el uso de las evaluaciones seguían siendo las mismas (el mecanismo induce pero no obliga a la utilización) y que tampoco se disponía de un programa concreto para atender los aspectos interinstitucionales e intergubernamentales. Además, al no ofrecer incentivos evidentes para su utilización, se podría estar presentando un círculo burocrático de simulación.

La falta de claridad sobre los niveles de utilización efectiva de resultados llevó también a la ASF (2010) a recomendar el establecimiento de mecanismos de seguimiento y disposición de información sobre los programas, metas y acciones de la política de desarrollo social que han sido corregidos, modificados, adicionados, reorientados, o suspendidos total o parcialmente.

Entre las generaciones de 2010-2011 y 2011-2012, el número de programas evaluados disminuyó pero aumentó el de ASM, con lo cual el promedio previo de 3.6 ASM por programa pasó a 5.7, lo que se asocia con la nueva aplicación de evaluaciones de Consistencia y Resultados en el segundo caso, que dio origen

al 77.1% del total de ASM, con 31.5% de los mismos vinculados al diseño del programa y 37.6% de alta prioridad.

Finalmente, en las evaluaciones de 2012-2013, 73 de los 260 ASM emitidos se relacionan con el diseño de programas y 72.7% proviene de EED. La disminución en el número de programas evaluados y de ASM se debió a que en el ejercicio fiscal de 2014 se efectuaron cambios sustantivos en la estructura programática de las dependencias y entidades, por lo que los programas que tuvieron evaluaciones externas en ejercicios anteriores pero que ya no aparecieron en el Decreto del PEF de 2014, decidieron no incorporar recomendaciones y/o hallazgos en el presente mecanismo.

En cuanto al cumplimiento de los ASM surgidos de las ECR 2007-2008, el informe de 2009 no presenta avances; en cambio, De la Garza (citada en Montiel, 2011) señalaba que 58% de los programas utilizaron las evaluaciones para mejorar su desempeño, siendo el DIF, la Semarnat y el Conacyt las instancias que no cumplieron plenamente con sus compromisos en materia de aspectos susceptibles de mejora.

De igual forma, el informe del año siguiente reporta un incremento de 60% en el número de programas que realizaron el seguimiento a las recomendaciones de las evaluaciones externas respecto del año anterior; pero tampoco informa sobre los avances realizados. Montiel (2011) mencionaba que alrededor del 45% de los programas reportaron acciones con un avance entre 80 y 100% respecto de sus EED 2008-2009, lo que mantenía bajo el nivel de cumplimiento, y que las acciones concluidas se referían principalmente a avances en la Matriz de Indicadores para Resultados.

A diferencia de los informes anteriores, los seguimientos de ASM subsiguientes ofrecen datos comparables, lo que permite concentrar su evolución en el Cuadro 38.

En el cuadro mencionado se observa que el promedio general de cumplimiento de los ASM informado ha tenido en general una tendencia creciente, pero muy variable de un año a otro (disminuciones de más de 10 e incrementos de casi 20 puntos porcentuales). Las instancias con mayor número de aspectos cumplidos coinciden también con las que presentan mayor número de ASM y de programas evaluados, por lo que eso no significa que su nivel de cumplimiento sea alto. Por ejemplo, en el penúltimo periodo informado la SEP, la CDI y la Semarnat sólo

CUADRO 38
Seguimiento de los ASM identificados desde 2010 a la fecha

Avances	Informe 2010-2011	Informe 2011-2012	Informe 2012-2013	Informe 2013-2014
Promedio general de avance (%)	69.4	73.9	62.6	81.9
Total de ASM cumplidos por los programas*	97	120	100	41
Total de ASM cumplidos en acciones*	329	354	405	71
Dependencias y entidades con más número de ASM cumplidos	CDI Semarnat Sedesol	SE CDI	SEP CDI Semarnat	SEP CDI Semarnat
Impactos de los ASM en acciones (%)	Adiciones 36	Adiciones 0	Adiciones 4	Adiciones 0
	Correcciones 22	Correcciones 71	Correcciones 64	Correcciones 72
	Modificaciones 18	Modificaciones 12	Modificaciones 4	Modificaciones 4
	Reorientaciones 22	Reorientaciones 15	Reorientaciones 28	Reorientaciones 22
	Suspensión 2	Suspensión 1	Suspensión 0	Suspensión 0

* Información de ASM totalmente cumplidos a septiembre de 2014, según el cuestionario respondido por el Coneval. Otras fuentes: Informes de Seguimiento de ASM de 2010-2011 a 2013-2014 (Coneval, 2011g, 2012e, 2013j y 2014b).

presentan avances del 42.0, 63.5, y 76.1% respectivamente de sus compromisos totales vencidos.

En cuanto al impacto de los ASM cumplidos, las llamadas adiciones, que incluyen fusiones con otros programas y reubicaciones en otras instancias, han tendido casi a desaparecer; las correcciones vinculadas con actividades, procesos e indicadores de la MIR concentran en promedio en los últimos tres años casi 70% del impacto; las modificaciones referidas a los bienes y servicios que ofrece el programa absorben un porcentaje muy pequeño (4% en los dos últimos años); las reorientaciones constituyen modificaciones sustantivas del propósito del programa y pueden implicar la realización de evaluaciones de impacto, por lo que su concreción es muy importante y se requeriría de una información mucho más detallada sobre los cambios y programas en que se han aplicado (promedio de 22% en el periodo analizado); y por último, las “suspensiones”

pueden implicar incluso la terminación total del programa, lo que sólo sucedió en tres casos de los primeros años, que el informe no identifica.

Algunos datos adicionales de los informes de estos cuatro años permiten comprender mejor los resultados de los procesos de evaluación. Así, el Informe de 2011-2012, en términos globales, reporta un cumplimiento pleno en 74 programas y en 56.6% de los ASM pendientes acumulados hasta 2011, y un promedio aproximado de 40% de avance en los restantes casos cuyo plazo ya había concluido. Por su parte, el informe de 2012-2013 muestra que los únicos tres organismos que lograron un pleno cumplimiento sólo tenían 1, 3 y 13 ASM asumidos. En este periodo se mantuvieron las altas tasas de cambios dirigidos simplemente a corregir actividades o procesos, pero aumentaron significativamente los vinculados a reorientaciones sustantivas del programa.

El último de los informes aparecido (2013-2014) reporta un cumplimiento pleno en el 65.4%² de los ASM pendientes acumulados, casi 9 puntos porcentuales más que dos años antes. Agrega que los aspectos concluidos se concentran en la prioridad media, preocupando que sólo alcance el 28% en el caso de los aspectos institucionales de prioridad alta. También se informa del avance de ASM en función de su año de terminación: 87% del 2012, 65% del 2013 y 52% del 2014. Un problema que se presentó fue que existieron ASM cancelados por las dependencias y entidades (Cuadro 39) debidos a cambios sustantivos de los programas presupuestarios (73 específicos y 10 institucionales) y otro, que siete dependencias o entidades no presentaron avances de 32 ASM durante este ciclo, resaltando la SEP con 12 casos (Cuadro 40), por lo que se decidió dar de baja a los ASM que no presentaban actividad (124 específicos y 4 institucionales). Con esta decisión se redujo considerablemente el incumplimiento reportado, sin que estos ASM cumplieran su función.

Ante la solicitud realizada al Coneval para que ejemplificara algunos casos concretos de utilización de los hallazgos y recomendaciones de sus evaluaciones presentó los siguientes:³

² El 55.9% en aspectos específicos y 70.6% en los institucionales.

³ Respuesta del Coneval al cuestionario elaborado para esta investigación.

CUADRO 39
ASM cancelados 2011-2014

Dependencia o entidad	Específicos 2011-2012	Específicos 2012-2013	Específicos 2013-2014	Institucionales 2013-2014	Total
Agrosemex	0	0	0	0	0
CDI	2	11	22	10	45
IMSS	0	0	0	0	0
Sedatu	0	0	0	0	0
SEP	0	0	35	0	35
Semarnat	0	1	0	0	1
Salud	0	0	2	0	2
Total	2	12	59	10	83

Fuente: reelaborado del Coneval (2014b).

CUADRO 40
ASM dados de baja 2012-2014

Dependencia o entidad	Específicos 2011-2012	Específicos 2012-2013	Institucionales 2013-2014	Específicos 2013-2014	Total
Agrosemex	1	0	0	0	1
CDI	0	0	0	0	0
IMSS	1	1	0	2	4
Sedatu	0	0	2	0	2
SEP	3	59	2	0	64
Semarnat	55	2	0	0	57
Salud	0	0	0	0	0
Total	60	62	4	2	128

Fuente: reelaborado del Coneval (2014b).

- a) Se establecieron topes a las transferencias de los beneficiarios con más propiedades del Programa de Apoyo al Ingreso Agropecuario para Vivir Mejor.
- b) Se extendió a zonas urbanas el Programa de Empleo Temporal debido a la insuficiencia en el crecimiento del empleo formal en las mismas.
- c) Educación Preescolar y Primaria para Niñ@s de Familias Jornaleras Agrícolas Migrantes, creó y fortaleció el Sistema Nacional de Control Escolar Migrante para contar con la información académica de los alumnos y evitar repetición y deserción.
- d) ProÁrbol simplificó el proceso de pago de los apoyos para hacer más eficiente su operación.
- e) Promajoven dio seguimiento a su Sistema de Información en todas las entidades federativas para incluir información sobre el tipo de escolaridad y la proporción de beneficiadas que transitan por las modalidades de escolaridad abierta, presencial o a distancia.
- f) El PDHO modificó su estrategia operativa en zonas urbanas al detectarse menores impactos que en zonas rurales.

Al respecto, Székely (2013) reconoce que la consolidación futura del Coneval depende de que los ejecutores de programas y políticas utilicen la información generada y el riesgo de que la definición y cumplimiento de los ASM se convierta en un proceso burocrático y heterogéneo, que no siempre es llevado a cabo por los responsables directos de los programas. Para mejorar esta situación recomienda que se ligue a incentivos concretos de reconocimiento, impacto en presupuestos, establecimiento de responsabilidades a servidores públicos en casos de falta de atención e incluso observaciones hacia la SFP si éstos no son considerados con la seriedad debida. Destaca, por ejemplo, que el ente rector de la política social (Sedesol) había atendido menos de la mitad de sus ASM a esa fecha, lo que muestra la necesidad de que las dependencias den mayor prioridad al tema y que el Coneval haga un mayor esfuerzo por acompañarlas en el proceso de elaboración de los documentos de trabajo que los retomen, mediante la asesoría y capacitación de los funcionarios. Incluso propone dar carácter vinculatorio a la atención de los aspectos susceptibles de mejora.

Si bien 96% de los funcionarios públicos encuestados por Székely expresó una opinión positiva respecto de las recomendaciones emitidas por el Coneval, en general, perciben los ASM institucionales, interinstitucionales, e

intergu-bernamentales como poco factibles, y por lo tanto, les prestan poca atención. Las principales razones expresadas de la falta de aplicación de las recomendaciones fueron las restricciones presupuestales (34.1%), seguidas de la falta de consistencia entre lo que el Coneval y otras instancias les exigen (34.1%), mientras el resto de las respuestas varía desde la no factibilidad de las recomendaciones y la normatividad que lo impide, hasta la falta de interés de los tomadores de decisiones.

En consecuencia, Székely informa sobre el impacto que el sistema de evaluación ha tenido en la Política Programática de Desarrollo Social en 2010 y 2011-2012 respectivamente, confirmando que la tendencia es a que disminuya la reorientación sustancial (rediseño) de los programas mientras aumenta la corrección de actividades y procesos (operación). A esto agrega que la percepción de los funcionarios sobre el grado de repercusión de estas recomendaciones en los programas ha sido variada. Los aspectos en que consideraron que tuvo algo o mucha repercusión fueron: metodología (35.4%), planeación (33.3%), resultados (29.2%), operación (22.9%), presupuesto (16.75%), impacto (16.7%) y un 2.1% señaló que han confundido al personal.

Como conclusión general del mecanismo analizado, se afirma que los porcentajes de cumplimiento general informados por el Coneval son considerables (81.9%) y crecieron drásticamente en el último año informado (2013-2014), en parte debido a la cancelación de 73 ASM específicos y 10 institucionales y la baja de 124 específicos y 4 institucionales, reportadas el año anterior. El cumplimiento se logra ágilmente en asuntos metodológicos relacionados con la MIR (metas, indicadores, etcétera) y en temas relativos a procesos de operación (ejemplos presentados por Cejudo y Maldonado, 2011), pero se ve más difícil cuando se trata de hacer un diagnóstico o un programa estratégico, una reformulación sustantiva de la política o una evaluación de impacto en la transmisión intergeneracional de la pobreza (caso de Oportunidades).⁴ También se dificultan aspectos que involucran a varias instancias o varios ámbitos

⁴ En 2008 Oportunidades llevó a cabo una evaluación de impacto del programa en áreas rurales, en materia de calidad de los servicios de salud, nutrición y educativos, y de la operación del programa y la atención de las familias beneficiadas, pero no en la transmisión intergeneracional de la pobreza, como se ha venido proponiendo.

de gobierno. Finalmente, la información disponible no permitió valorar si existe relación entre el peso económico de llevar a cabo un ASM y su nivel de cumplimiento.

Otro de los mecanismos ideados por el Coneval para realizar la promoción de la utilización de las recomendaciones ha sido el reconocimiento de Buenas Prácticas en el Uso de los Resultados de Monitoreo y Evaluación (no en el desempeño de los programas) desde 2009, los que se sintetizan en el Cuadro 41. En 2013 también se entregó un reconocimiento similar a las entidades federales que muestran avances de monitoreo y evaluación.⁵

Como puede verse, en los últimos años ya no se relacionan siempre con la utilización de resultados de monitoreo y evaluación. Se desconoce el impacto logrado por este tipo de reconocimientos, que no ha sido evaluado.

Una alternativa al problema general de poca utilización de las recomendaciones en los programas federales podría ser la intentada por Evalúa DF, que será revisada más adelante, en el sentido de que éstas tengan carácter vinculatorio. Esta posibilidad es analizada por Székely (2013) para el nivel federal como una discusión compleja, donde 9.59% de sus encuestados señaló la falta de este carácter como una debilidad del sistema. El autor afirma que desde el punto de vista del evaluador propiciaría un mayor uso de la evidencia, pero desde la perspectiva de los operadores de programas las evaluaciones no necesariamente captan el contexto socioeconómico, cultural y político específico en que la operación se desenvuelve, lo que puede provocar recomendaciones no viables e incluso no deseables que afectarían negativamente al programa. Además, aunque el responsable del programa tenga la voluntad de instrumentar un cambio, puede suceder que no cuente con los recursos presupuestales necesarios. En consecuencia, imponer sanciones al incumplimiento en este contexto podría generar incentivos negativos. El reto consiste en presentar los resultados de manera tal que resulten relevantes, comprensibles y fácilmente utilizables, y que los responsables de las acciones cuenten con los incentivos adecuados para llevarlas a la práctica.

⁵ Evalúa DF fue reconocido entre éstos.

CUADRO 41

Reconocimiento de buenas prácticas en el uso de los resultados de monitoreo y evaluación

Año	Reconocimientos y menciones honoríficas	Casos de la Sedesol y Evalúa DF
2009-2010	12	<ul style="list-style-type: none"> • Fonart, por su matriz de indicadores. • CDI, por su sistema de seguimiento de indicadores • Oportunidades, por su matriz de indicadores y también por su proceso de evaluación continua
2010-2011	9	<ul style="list-style-type: none"> • Sedesol por encuesta sobre percepción de inseguridad, conductas de riesgo y participación social en espacios públicos • Sedesol por implementación de un proceso participativo de la evaluación • CDI por construcción de un sistema informático para simplificar el proceso de control y seguimiento de las recomendaciones de evaluaciones externas
2011-2012	8	<ul style="list-style-type: none"> • Sedesol por la elaboración de diagnósticos de los programas sociales • Sedesol por agenda de evaluación 2007-2012 del Programa de Estancias Infantiles • Sedesol por coordinación institucional para el diseño e implementación de la evaluación de impacto de Hábitat 2008-2012
2012-2013	1 reconocimiento y 4 menciones tanto en Monitoreo como en Evaluación	<ul style="list-style-type: none"> • Mención honorífica a Oportunidades por su modelo de seguimiento operativo • Mención honorífica a Sedesol por las metaevaluaciones de los programas sociales
2013-2014	3 reconocimientos por innovación 4 por uso de la evaluación y desarrollo de capacidades 4 por su sistema de indicadores	<ul style="list-style-type: none"> • Diseño de Evaluación de Impacto de IMSS-Prospera • Uso de las evaluaciones para el rediseño de Prospera • Uso de las evaluaciones para el rediseño programático de CDI • Sedesol por la elaboración de las Fichas de Monitoreo y Evaluación 2013-2014 • Sedesol por el diseño e implementación de políticas sociales mediante el uso de indicadores de pobreza multidimensional: CNCH y Fondo de Aportaciones de Infraestructura Social

Fuente: [http://www.coneval.gob.mx/Evaluacion/Paginas/Buenas_Practicas], fecha de consulta: 30 de octubre de 2014.

Para resolver el problema, Székely recomienda en general una mayor participación y comprensión de los programas en el proceso de evaluación: presentar el ejercicio de evaluación como un instrumento para la mejora, brindar mayores oportunidades de capacitación y asesoría técnica a las dependencias, garantizar que las evaluaciones se enfoquen en elementos de interés para los responsables de los programas y acciones sociales, reforzar la comunicación entre el evaluador, las dependencias y el Consejo durante el desarrollo de la evaluación, realizar reuniones de trabajo intermedias en donde los evaluadores presenten resultados preliminares, revisar la factibilidad de las recomendaciones emitidas, fortalecer los mecanismos para que los resultados de las evaluaciones sean revisados y considerados por los funcionarios de mayor rango jerárquico de los programas de tal forma que se propicie su utilización, implementar estrategias de mejora de la comunicación con otras secretarías (especialmente SHCP y SFP) y difusión de los productos hacia un público más amplio, entre otras. Con esto recoge una buena parte de las propuestas surgidas en la revisión teórica realizada en el capítulo previo.

Concretamente, respecto de los ASM, los participantes en grupos focales y los encuestados en el trabajo de Székely señalaron la tendencia a aceptar más los específicos, puesto que los institucionales, interinstitucionales e intergubernamentales requieren de mayor coordinación entre diversas áreas de las dependencias o entre dependencias y niveles de gobierno, manifestaron la conveniencia de que el Coneval se involucre más en el proceso de determinación e implementación de los ASM, de los que prácticamente sólo hace un seguimiento cuantitativo, e insistieron en que las recomendaciones de los evaluadores externos resulten viables y adecuadas a los tiempos fiscales. Además, algunos de ellos comentaron que no tienen acceso al sistema de seguimiento a los ASM perteneciente al Coneval, por lo que se ven forzados a duplicar el esfuerzo en sus dependencias. Székely reconoce que existe un potencial conflicto de interés, debido a que son las mismas unidades responsables quienes los determinan y sugiere que en la determinación de las recomendaciones que se convertirán en ASM participen también los responsables de la programación y presupuesto a nivel de cada dependencia, así como el Órgano Interno de Control como ente verificador y que se valore la conveniencia de otorgar al Coneval más atribuciones en la determinación y el seguimiento de ASM, incluyendo la posibilidad de que sus recomendaciones resulten vinculantes.

- Algunas experiencias exitosas de utilización instrumental

En paralelo con las evoluciones relatadas a nivel federal, como ya se mencionó, Cejudo y Maldonado (2011) presentan las experiencias que consideran especialmente exitosas, aunque atípicas y con desafíos pendientes, en materia de recuperación de las recomendaciones, las que se sintetizan en el Cuadro 42.

CUADRO 42
Mejoras instrumentadas por evaluación y programa

Programa	Evaluaciones	Mejoras instrumentadas
Sistema Integral de Calidad en Salud	<ul style="list-style-type: none"> • Diseño 2008 • EED 2009 y 2010 • Procesos 2010 	<ul style="list-style-type: none"> • Instalación de comités y gestores de calidad en hospitales y unidades médicas • Medición de la cultura de la calidad • Acreditación y capacitación en temas de calidad • Difusión del programa
Expedición de Pasaportes y Servicios Consulares	<ul style="list-style-type: none"> • Evaluación 2010, tipo no especificada 	<ul style="list-style-type: none"> • Establecimiento del Programa de Citas Mexitel • Primera fase del Sistema Integral de Registro de Mexicanos en el Extranjero • Cursos de capacitación en los Consulados de Canadá, Texas y California • Actualización de dos capítulos de la nueva Guía Consular • Reestructuración de la MIR
Atención a la Demanda de Educación para Adultos y Modelo de Educación para la Vida y el Trabajo	<ul style="list-style-type: none"> • ECR 2008 • Procesos 2009 	<ul style="list-style-type: none"> • Valoración de 33,969 asesores (52.6% del total) • 44% de avance en la valoración y revaloración diagnóstica.
ProArbol. Programa de Pago por Servicios Ambientales	<ul style="list-style-type: none"> • EED 2010-2011 • ECR 2011-2012 	<ul style="list-style-type: none"> • Línea basal para la futura comparación del desempeño de los indicadores • Implementación de un sistema de monitoreo para evaluar el impacto del programa.

continúa...

CUADRO 42

Mejoras instrumentadas por evaluación y programa

Programa	Evaluaciones	Mejoras instrumentadas
Escuelas de Tiempo Completo	• Diseño 2008	<ul style="list-style-type: none"> • Diseño del programa de seguimiento de la implementación de la propuesta (Simaetic) • Campus virtual de asesoría a coordinadores y responsables del programa en las entidades federativas • Desarrollo de competencias y habilidades para el registro de información • Diseño de un trabajo de campo en 125 escuelas para revisar los reportes de resultados
Escuelas de Calidad	• Complementaria 2008 • EED 2009	<ul style="list-style-type: none"> • Publicación y distribución del Plan Estratégico de Transformación Escolar Simplificado • Redacción y distribución del Módulo “Orientaciones para activar la participación social en las escuelas de educación básica” • Conferencia e intercambios para promover el uso de tecnologías de la información • Encuentro de Seguimiento y Evaluación de la Red de Formadores • Fichas de información estadística por entidades para mejorar la focalización del programa
Fondo para el Apoyo a Proyectos Productivos en Núcleos Agrarios	• Complementaria de Fin y Propósito 2009	<ul style="list-style-type: none"> • Supervisión previa de la veracidad de la información de las solicitudes de apoyo y reasignación de recursos en su caso
Mujer en el Sector Agrario	• Complementaria de Fin y Propósito 2009	<ul style="list-style-type: none"> • Supervisión previa de la veracidad de la información de las solicitudes de apoyo y reasignación de recursos en su caso • Indicador de igualdad de trato entre hombres y mujeres
Becas de Posgrado y otras Modalidades de Apoyo a la Calidad Educativa	• Impacto, BM, 2007 • EED 2008 y 2009	<ul style="list-style-type: none"> • Alineamiento del Conacyt con las recomendaciones dirigidas a contar con capital humano altamente calificado y atraer estudiantes del extranjero para realizar estudios de posgrado en el país (no especifica las acciones concretas realizadas)
Joven Emprendedor Rural y Fondo de Tierras	• EED 2010-2011 • Complementaria 2013	<ul style="list-style-type: none"> • Estudio de FAO para identificar los factores que determinan el éxito de una agroempresa • Capacitación para beneficiarios con Financiera Rural • Información generada mediante supervisión sobre número de jóvenes beneficiarios, sobrevivencia empresarial y respeto ambiental

Fuente: elaboración propia a partir de Cejudo y Maldonado (2011).

- Utilización presupuestal

La utilización presupuestal se refiere a la asignación anual de recursos a las políticas y programas públicos en marcha, que debería considerar los resultados de las mismas, tratando de consolidar aquellas que resultan más efectivas y de mayor impacto, así como de reforzar las que demuestran que sus avances limitados se deben a insuficiencia presupuestaria. Este tipo de utilización también incluye elementos presentes de la utilización persuasiva.

Una forma de verificar si las evaluaciones realizadas en México han incidido en el complejo proceso técnico-político de asignación de presupuesto consiste en contrastar este último con los resultados alcanzados.⁶ En el caso federal, el análisis realizado a la evolución del gasto ejercido en 2007-2012 (Cuadro 43) muestra que, en su conjunto, éste alcanzó un incremento del 73%.⁷ Sin embargo, se observan notables variaciones por programas, siendo pocos los que se acercan a la media (Fondo Nacional de las Artesanías y 3X1 Migrantes). En cambio aparecen programas que disminuyen notablemente su presupuesto (Abasto Social de Leche, Abasto Rural y Oportunidades Productivas), otros que se mantienen prácticamente sin cambios (Servicios a Grupos con Necesidades Especiales y Apoyo a las Instancias de Mujeres en las Entidades Federativas), algunos que presentan importantes incrementos (Estancias Infantiles, 70 y Más, y Empleo Temporal), y uno con un aumento notable (Apoyo Alimenticio). A continuación se analizan los casos más extremos (Coneval, 2011b y 2013m).

En 2011 Opciones Productivas no había tenido evaluaciones de impacto que justificaran cambios presupuestales relevantes. Sin embargo, la reducción de su presupuesto inició en 2010, se agudizó en el siguiente año y en 2012 alcanzó casi la mitad de su presupuesto de 2010. Según sus evaluadores, su disminución presupuestal no se debía a una reducción de las necesidades de ingreso que

⁶ Una primera versión de este análisis presupuestal se encuentra en Cardozo (2013) con datos de 2008-2010, que aun cuando también se basó en datos del Coneval presenta diferencias con los actuales pero marca tendencias muy similares [http://www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/evaluaciones/especificas/edd_2010_2011/pag_princ_edd_2010_2011.es.do].

⁷ El Coneval no aclara si la información está en precios corrientes, pero se supone que así sea dada la importancia de los incrementos reportados.

generaban sus proyectos, sino más bien a un ajuste técnico-metodológico de su población potencial, definida en función del presupuesto disponible y no a la inversa, situación que siguió argumentándose en 2012-2013. En 2008 su población potencial era de 20 millones de personas y la población objetivo de 10 millones, atendiéndose sólo el 0.50% de la misma. Debido al proceso sufrido, en 2009 los dos tipos de población se redujeron a 5 millones, mientras que la atención en 2012 se limitó al 0.28 por ciento.

Abasto Social de Leche, según sus evaluaciones de impacto (2004, 2006 y 2009), sólo había mostrado avances claros en materia de leche fortificada y en ciertas edades. En 2006 no se encontró impacto en el desarrollo cognitivo de los niños; tampoco en 2009 se identificó impacto en el desempeño escolar; en cambio, mejoró el desarrollo intelectual pero éste se explicaba por otra causa: el nivel socioeconómico (Coneval, 2011c). En 2012 sus evaluadores cuestionaron la calidad de dichas evaluaciones y sostuvieron, con aportes de la ECR 2011-2012, que el desempeño del programa seguía siendo poco evidente, si bien tenía una cobertura casi total de su población objetivo y la satisfacción de sus beneficiarios alcanzó el 9.12 (en una escala de 0 a 10). En este caso algunos de los resultados de las evaluaciones que han sido mencionados son consistentes con la tendencia presupuestal a la baja, que continuó hasta 2012.

Abasto Rural es un programa que, a pesar de que su metodología ha recibido críticas, atiende a toda su población objetivo y lo hace en forma creciente. El aumento de su crecimiento coincide paradójicamente con la reducción en 33 puntos porcentuales de su presupuesto en 2012. No obstante, no cuenta con una evaluación del impacto.

Apoyo Alimentario fue el programa con un aumento más notable, especialmente en 2010, aunque el monto absoluto de su presupuesto estaba por debajo de Desarrollo de Zonas Prioritarias, 70 y Más, y, muy especialmente, Oportunidades. En 2009-2010 se fusionó con el Programa de Apoyo Alimentario en Zonas de Atención Prioritaria y pasó de ser responsabilidad de Diconsa a la Coordinación de Oportunidades, programa con el que se asimiló, cambiando su población objetivo y condicionando su apoyo a la asistencia a unidades de salud,⁸ etcétera.

⁸ A pesar de que opera en muchas localidades sin estos servicios ni los educativos.

Si bien como en muchos de los programas sociales existe una discusión importante sobre la cuantificación de sus poblaciones y cobertura, en este caso incluso entre Sedesol y el programa, en 2010 la población atendida se habría más que duplicado para mantenerse luego constante, lo que explicaría la evolución presupuestal.

Estancias Infantiles presentó su mayor incremento presupuestal entre 2007 y 2008 por tratarse de un programa reciente, muy impulsado desde el poder Ejecutivo, que iba ampliando su cobertura. Entre 2008 y 2012 el aumento de sus recursos es importante (55%) pero incluso menor al promedio general.

De similar forma, el presupuesto de 70 y Más, en el periodo 2007-2012, fue más que duplicado, esto se explica fundamentalmente por los incrementos obtenidos en 2009, año en el que también duplicó su población potencial, pero mantuvo un crecimiento muy leve de su población objetivo y atendida, generando una brecha de más de dos millones de personas sin atención, y en 2012, fecha en que el aumento de la población objetivo fue más fuerte que el de la potencial y de la atendida, con lo cual la brecha mencionada apenas ha disminuido. Se trata entonces de un caso en el cual los aumentos presupuestales buscarían promover una cobertura que el programa no logra, en lugar de premiar su desempeño.

Empleo Temporal tuvo un aumento importante de su población atendida entre 2007 y 2010 (pasó de 240 000 a 582 044 personas). El 71% de sus beneficiados manifestaban mejoras en sus condiciones de vida y en 79% de sus ingresos. En consecuencia, el beneficio percibido y la creciente necesidad de empleo justificaron el notable aumento de su presupuesto en 2009, pero en 2010 éste se redujo casi a la mitad para mantenerse luego prácticamente estancado, a pesar de que en 2012 la población potencial es de casi el quíntuple de la población objetivo. La población atendida creció lentamente hasta 2011 y cubría 100% de la población objetivo; sin embargo, en 2012 disminuyó significativamente para cubrir sólo 76% de esta última. Como en otros casos, se reportan diferencias de cálculo de la población objetivo entre la Sedesol y la Semarnat y el programa también carece de una evaluación de impacto, a pesar de su antigüedad. En todo caso, serían las reducciones presupuestarias iniciadas en 2010 las que deben estar provocando una mayor ineficacia del programa en 2012 y no a la inversa.

En resumen, a pesar de que el Coneval genera y difunde información para el proceso presupuestario en función de las evaluaciones realizadas, la Cámara de

CUADRO 43
Presupuesto por programas de la Sedesol 2007-2012
(millones de pesos*)

Programa	2007	2008	2009	2010	2011	2012	Aumento porcentual 2007-2012
DH Oportunidades	44 276	48 568	50 936	61 283	61 113	66 092	49
70 y Más	7 448	11 260	13 912	13 912	13 506	17 693	138
Des. Zonas Prioritarias	0	4 449	8 262	7 847	6 118	6 189	39 (inicio.2008)
Apoyo Alimentario	356	486	1 350	3 817	3 847	4 053	1 038
Estancias Infantiles	908	2 020	2 671	2 619	2 668	3 125	244
Abasto Rural	2 906	2 367	2 026	2 139	2 693	1 801	-38
Adq. Leche Nacional	0	0	1 402	1 519	1 810	1 689	20 (inicio.2009)
Empleo Temporal	744	576	2 912	1 566	1 546	1 587	113
Abasto Social Leche	2 174	3 237	1 459	1 260	1 229	1 050	-52
3X1 Migrantes	304	581	581	584	569	489	61
Coinversión Social	338	322	377	441	367	427	26
Opciones Productivas	1 100	1 360	1 048	816	498	413	-62
At. Jornaleros Agrícolas	188	174	239	239	212	271	44
Apoyo Inst. Mujeres EF	230	229	205	205	207	232	0
Serv. Grupos Nec. Esp.	0	228	215	229	230	229	0 (inicio.2008)
Fondo Nac. Artesanías	89	116	90	96	78	144	62
Subsidios Progr. Jóvenes	0	0	46	65	130	62	35 (inicio.2009)
Total	61 061	75 973	87 731	98 637	96 821	105 546	73

* Se supone en precios corrientes, los documentos no lo especifican.
Fuente: elaboración propia a partir de las Fichas de Monitoreo de Programas Sociales 2012-2013 (Coneval, 2013m).

Diputados ha continuado tomando sus decisiones con una casi nula utilización de los resultados de las mismas, a menudo desconocidas por sus integrantes o confundidas con los trabajos de la ASF (Cardozo, 2009). El análisis efectuado a los casos más relevantes de incrementos o reducciones presupuestales mostró poca relación con los resultados de las evaluaciones realizadas a los programas federales, salvo en el caso de Apoyo Alimentario y, hasta cierto punto, de Abasto Social de Leche. En dos de los casos la disminución no fundamentada del presupuesto parecería estar provocando la ineficacia del programa (Opciones Productivas y Empleo Temporal). En 70 y Más el presupuesto se incrementa a pesar de que la eficacia no lo hace y en Abasto Rural la cobertura aumenta a pesar de la reducción de los recursos, lo que sugiere la importancia de conocer su impacto real en el problema alimentario. La ausencia de evaluaciones de impacto surge como uno de los *talones de Aquiles* de este proceso, pues sólo Abasto Social de Leche las dispone pero se encuentran desactualizadas y metodológicamente cuestionadas por evaluaciones más recientes.

Finalmente, el problema de la disonancia entre las evaluaciones y las asignaciones presupuestales también fue considerado por Székely (2013), quien señala que, salvo algunas excepciones, no se observa que las evaluaciones incidan en las decisiones de reasignar recursos, evitar duplicaciones, o lograr consolidaciones. De hecho, percibe una relación poco clara entre el PAE y el presupuesto debido a que las evaluaciones se concluyen normalmente una vez que los presupuestos propuestos en el PEF del año subsecuente ya han sido determinados. Incluso se dan casos en los cuales se detectan elementos de vulnerabilidad o diseño imperfecto de los programas, pero estos reciben un presupuesto similar o mayor en el siguiente ejercicio a pesar de no realizar los cambios que la evaluación sugiere. En particular, no existe claridad en cuanto a las consecuencias de las EED y las ECR en el presupuesto de ejercicios posteriores. Consistentemente con este argumento, sólo 16.75% de los funcionarios que Székely entrevistó consideró que las recomendaciones tienen impacto presupuestal. Muchos señalaron que diversos programas que no han obtenido los resultados esperados o no son progresivos, no han sido eliminados debido fundamentalmente a cuestiones políticas y a que el Coneval se limita a la presentación de la información sin mayores posibilidades de influir en el proceso de asignación de recursos.

El autor agrega que la Dirección General de Programación y Presupuesto de la SHCP, que debe incidir en el presupuesto de los programas y dependencias, actualmente se enfoca más en el ejercicio del gasto público que en los indicadores de la MIR o en los resultados del SED, a los que en 2006-2012 atribuía más importancia, lo que induce a percibir que estos procesos tienden a transformarse en requisitos burocráticos de poca utilidad.

- Utilización conceptual

Por último, la utilización conceptual, como ya fue definida, se refiere a modificar una idea o actitud a través de un proceso de aprendizaje generado a partir de la información ofrecida por la evaluación, que incluso podría ser utilizada de manera instrumental en el futuro. Evidentemente, esta forma de utilización puede ser muy importante pero es mucho más difícil detectar al conjunto de actores que afecta y, más aún, operativizar su medición. Probablemente, la manera más sencilla sería aplicar una encuesta a los principales actores que podrían haber realizado este tipo de utilización, pero no se tiene conocimiento que se haya efectuado alguna.

Entre los posibles actores impactados conceptualmente habría que incluir a los funcionarios y legisladores vinculados con la utilización instrumental y presupuestal, ya revisada. Pero también se debería agregar a los propios evaluadores, investigadores universitarios del tema, funcionarios que, si bien no toman decisiones instrumentales, se vinculan con la evaluación (diseñan RO y TR, analizan su cumplimiento, etcétera), y a todos los ciudadanos que lean los informes y escuchen o participen en su debate.

Un importante prerequisite para que este aprendizaje conceptual se pueda producir es el acceso a los resultados de las evaluaciones. En ese sentido se han producido avances muy importantes desde la aprobación de la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental en 2002, que permitió la obtención de informes a solicitud expresa, y la correspondiente a la LGDS al siguiente año, que exigió la publicación de los mismos, lo que se ha traducido en una muy amplia difusión de los informes en las páginas institucionales del Coneval y de las dependencias y entidades responsables de programas.

Al respecto, en 2011 y 2012, el Coneval (2012g) elaboró un índice de difusión de las evaluaciones; el primero reporta sobre las EED 2010-2011 y alcanza 100%, mientras el segundo refiere a las ECR 2011-2012 y se cuantifica en 92% debido a que dos de las evaluaciones realizadas, que corresponden a la SEP, no se publicaron. El documento señalaba como las instancias más reticentes: Comisión Nacional de Vivienda (Conavi), SS, Inmujeres y Conacyt y como el aspecto menos difundido, el correspondiente al costo de las evaluaciones (88%).

En años posteriores dicho índice dejó de publicarse, probablemente por considerar que la difusión estaba plenamente institucionalizada; sin embargo, recientemente resulta muy preocupante la entrada en vigencia de un requisito de autenticación (nombre y contraseña) para acceder a parte de la información de la página web del Coneval (ejemplo, en algunos ítems del tema Evaluación de la Política Social de su menú), lo que significa un retroceso a su previo acceso total.

En ausencia de una medición subjetiva del aprendizaje conceptual realizado, sólo es posible remitirnos al análisis de los productos generados, suponiendo que los avances identificados responden, al menos parcialmente,⁹ a un mejor manejo conceptual. Esto podría estar detrás de la adaptación del marco normativo de la evaluación, de los cambios registrados en la organización del sistema de evaluaciones externas, de la necesidad de incluir la evaluación del diseño de los programas a partir de un Marco Lógico, de la mayor precisión identificada en RO, TR y lineamientos de evaluación, entre otros.

En síntesis, la utilización conceptual constituye un tema totalmente abierto a futuras investigaciones, del que poco se puede decir en este momento.

- Desarrollo social

En una revisión general del caso de Sedesol en los últimos tres ciclos presupuestales se comprometió a realizar 194 ASM, con el más alto nivel del componente institucional (56%), lo que dificulta su cumplimiento al requerir el acuerdo entre dos o más áreas de la dependencia (Cuadro 44). Tal vez esta es parte de la razón que la ha llevado a reducir su porcentaje de ASM terminados

⁹ Otro factor explicativo de esta mejoría podría ser la capacitación.

a sólo 22% en el último ciclo, mientras en el anterior había alcanzado el 66%. Además en el último ciclo aumentó su número de ASM específicos, ubicándola en segundo lugar, después de la SEP. A pesar de esta situación no presentó cancelaciones ni bajas de ASM (cuadros 39 y 40).

Por otra parte, de las 206 recomendaciones recibidas en 2012-2013, la Sedesol sólo retomó 58 (28%), lo que la coloca en el nivel más bajo, quedando pendiente su cumplimiento hasta marzo de 2015. También ha recibido ASM interinstitucionales, uno de éstos repetido en los últimos ciclos, lo que indica incumplimiento (utilizar indicadores de desarrollo infantil en el programa de Estancias Infantiles).

CUADRO 44
Número de ASM por tipo en la Sedesol

Tipo de ASM	Ciclo 2011-2012	Ciclo 2012-2013	Ciclo 2013-2014	Total de los tres ciclos
Específicos	25	27	34	86
Institucionales	39	45	24	108
Total	64	72	58	194

Fuente: elaboración propia con información del Coneval (2012e, 2013j y 2014b).

Ante la imposibilidad de hacer un seguimiento de cada uno de los programas del sector, se presenta un acercamiento más concreto en el caso del programa social más importante del periodo en estudio (Oportunidades, actualmente Prospera).¹⁰ En el Cuadro 45 se ha intentado concentrar la información de los primeros años en que las EED permitieron hacer un seguimiento más estricto; sin embargo, al pasar los años la situación se complejiza porque empiezan a acumularse ASM no cumplidos de años anteriores y no siempre se dispone de un informe que actualice los datos. Por dicha razón, el cuadro se complementa con información cualitativa que permita una mejor comprensión del proceso.

¹⁰ Un primer análisis de los ASM de este programa se publicó como parte del artículo de Cardozo (2013).

La ECR de Oportunidades 2007-2008 (PDHO, 2008) realizó un conjunto de 14 recomendaciones de importancia sustantiva, de las que se destaca sintéticamente:

- Elaborar un diagnóstico actualizado y adecuado de la problemática que justifica el programa.
- Revisar la definición de las poblaciones potencial y objetivo y los métodos y criterios para identificarla y cuantificarla.
- Evaluar el impacto del programa en la transmisión intergeneracional de la pobreza.
- Elaborar un programa institucional estratégico de mediano y largo plazo.
- Valorar la conveniencia y factibilidad de que el Programa de Abasto Social de Leche y Oportunidades no sean excluyentes, sino que se complementen.
- Llevar Oportunidades u otros programas similares a las localidades más marginadas, pequeñas y dispersas, que son las que más los necesitan.
- Reducir gradualmente las becas generalizadas en primaria, focalizando más en los casos de mayor impacto, así como incrementar los apoyos en niveles subsecuentes.
- Diferenciar los montos de los apoyos, especialmente por tamaño de la familia, zonas rurales, oportunidades de trabajo y costos de accesibilidad.

CUADRO 45
Definición y seguimiento de ASM de Oportunidades 2007-2010

	2007-2008	2008-2009	2009-2010	2010-2011
ASM del año	ECR: 2 específicos 2 institucionales	EED: 8 Evaluación Impacto: 17 específicos 10 interinstitucionales/ intergubernamentales	5	8
Cumplimiento ASM años anteriores	--	2 específicos cumplidos 2 institucionales Retrasados	ND	17 específicos cumplidos

Fuente: Informe de Evaluaciones Específicas de Desempeño 2008-2011, tomado de Cardozo (2013).

Según el informe de la EED 2009-2010 (PDHO, 2010), que da seguimiento de los ASM derivados de las evaluaciones externas 2007-2008 las recomendaciones revisadas se habrían concretado en seis ASM pero en la EED de 2009-2010, sólo se retoman de ese periodo dos ASM específicos y dos institucionales, que dejan fuera importantes asuntos planteados en las recomendaciones (redefinición de poblaciones, evaluación de impacto, etcétera). Los primeros dos recuperan: la elaboración de un documento para sistematizar los mecanismos para el establecimiento de indicadores y metas; y, el fortalecimiento de algunos indicadores de la MIR en temas relacionados con el mejoramiento de la salud, el tránsito a la educación media superior y la recepción de apoyos alimentarios por parte de los beneficiarios. Los segundos incluyen: la elaboración de un diagnóstico actualizado y adecuado de la problemática que sustenta la razón de ser del programa, y la preparación de un Programa Estratégico. Estos últimos dos puntos son los únicos recuperados del listado previo de recomendaciones.

En el seguimiento de estos cuatro ASM se expresa que los primeros dos se cumplieron plenamente en la fecha comprometida; mientras que los dos institucionales habían sufrido retrasos: el primero se concluyó seis meses después de lo acordado, y habría dado lugar al documento “Movilidad social intergeneracional de los jóvenes beneficiarios de Oportunidades provenientes de hogares en zonas rurales”; mientras que el segundo sólo presentaba un avance del 50% al 31 de diciembre de 2009.

La EED de 2008-2009 (PDHO, 2009) reconoce que ha habido una “reflexión” sobre los ASM señalados en 2007-2008 en un Documento Institucional y otro Documento de Trabajo del programa Oportunidades; sin embargo repite, aunque en forma más sintética, las primeras cinco recomendaciones de la evaluación previa, ello hace suponer que no percibió un avance contundente en ellas.

Los ASM de las evaluaciones de 2008-2009 reaparecen en 2010-2011 y habrían sido dos específicos, dos institucionales, tres interinstitucionales y uno intergubernamental.¹¹ Se informa que fueron atendidos los específicos; seguían en proceso la elaboración del plan estratégico, si bien se contaba con una versión preliminar; y se carecía de información sobre los avances de los dos últimos tipos de aspectos.

¹¹ Evidentemente no hay una clara coincidencia de lo que se va reportando cada año.

En 2008-2009 también se realizó la evaluación de impacto de Oportunidades, que se tradujo en varios tomos de informes y 220 recomendaciones (su metaevaluación está incluida en la parte II de este libro). De éstas surgirían 27 ASM: 10 interinstitucionales o intergubernamentales y 17 específicos, con amplia incidencia referida a la calidad de los servicios de salud (tomo II) o a la difusión de la información y sin impacto en materia de diseño. Muchas se descartaron por considerar que ya estaban en proceso o por requerir de mayores recursos presupuestarios por lo que se prefirió dedicar sus recursos a ampliar la cobertura del programa.

Nuevamente la evaluación de desempeño 2009-2010 define otros ASM, aunque mucho menos claros y sustantivos que los anteriores, algunos redactados como recomendaciones, pero otros en términos de objetivos o simplemente de problemas identificados, como los siguientes:

- La ausencia de mecanismos formales para la interacción entre personal de Oportunidades y del Sector Salud a nivel local impide el flujo de información sobre la situación de los beneficiarios.
- Ampliar la información a la población beneficiaria sobre la disponibilidad de servicios de salud de segundo y tercer nivel.
- Identificar áreas de oportunidad para expandir las becas educativas del Programa.
- Las familias indígenas están en condiciones de desventaja y son más propicias a la exclusión cuando los encuestadores y el personal del Programa no hablan la lengua indígena de la zona.
- Los incentivos para lograr la participación de las vocales son muy limitados.

En 2010-2011 (PDHO, 2011), como seguimiento de los ASM de las evaluaciones externas de 2008-2009, el informe menciona como atendidos los 17 específicos y sin información de los 10 interinstitucionales e intergubernamentales. En cambio, el documento de Valoración del Desempeño de los Programas (Coneval, 2011d), reporta un total cumplimiento de las recomendaciones recibidas. La Sedesol informa del nivel de cumplimiento de cinco ASM, pero en todos los casos se trata sólo de elaboración de documentos y materiales informativos, que no incluyen cambios sustantivos en el funcionamiento y, menos aún, en el diseño del programa.

También se informó que se realizan dos evaluaciones de los efectos de mediano plazo de Oportunidades en zonas urbanas, una en educación y otra en salud y nutrición, las cuales están por concluirse, y que las RO de 2011 redefinían la población objetivo de acuerdo con la nueva metodología de medición multidimensional de la pobreza del Coneval. Existe también un documento de la Sedesol (2012b) en que se detectan y clasifican los ASM provenientes de las EED 2010-2011, que además tiene la virtud de explicar la razón de las recomendaciones descartadas, información muy difícil de obtener. Refiere la aceptación de ocho recomendaciones transformadas en ASM, la mayoría específicas (6), de prioridad media (6) y de factibilidad media (7). Respecto de las recomendaciones no atendidas (7 en total), éstas son dos en el caso específico de Oportunidades: *a*) aclarar en las RO que su población potencial y objetivo es la pobreza extrema y estructural, lo que fue rechazado por manejar conceptos no contemplados en la Metodología de Medición Multidimensional de la Pobreza a la que se ajusta el programa; y *b*) modificar el indicador “Porcentaje de cobertura de familias beneficiadas” por considerar que no mide la cobertura real del programa respecto del total de población a la que se dirige sino sólo en relación con la que se planea atender en el año. A esto se respondió que ya se había atendido, aclarando que el indicador se refiere a la población objetivo definida para el ejercicio fiscal correspondiente.

A partir de 2011-2012 la información del cumplimiento se reduce notablemente. La ECR realizada al programa en 2011-2012 no hace referencia a los avances de los ASM previos. Los ASM específicos, comprometidos para ser realizados en 2012, fueron seis.¹² No hay un documento relativo a aspectos institucionales.

¹² Los ASM incluyeron: analizar la posibilidad de incorporar al programa en la Encuesta Nacional de Ocupaciones y Empleo (ENOE), generar relaciones interinstitucionales con dependencias con experiencia en atender el rezago social de la población indígena, continuar la estrategia de comunicación del programa de Apoyo a Proyectos de Comunicación Indígena (PCI) y el uso de los materiales, e impulsar su mejora, fortalecer la capacitación del personal de salud y comunitario conforme a la estrategia del PCI, tratar de garantizar un espacio con las condiciones básicas para desarrollar los talleres del PCI, e impulsar el PCI en la modalidad de las Caravanas de la Salud y/o equipos de salud itinerantes.

En 2012-2013 se llevó a cabo una EED que dio lugar a nueve ASM¹³ (7 específicos y dos institucionales) a realizar en 2013-2014 y al seguimiento de un ASM específico y otro institucional, comprometidos en años anteriores. El primero es uno de los seis referidos el año anterior, relativo a incorporar el programa en la ENOE para el cual se informa que ya se tuvo comunicación con el INEGI para concretarlo. El ASM institucional revisado se refiere a la elaboración de un diagnóstico integral de cobertura y recursos destinados por los programas de transferencia de ingresos y alimentarios a la población pobre. Se reporta que éste ya se llevó a cabo, que ofrece buenos elementos de análisis, especialmente respecto de la medición de carencias de alimentación, pero que no comparó los programas en función de su costo/efectividad y su impacto, diagnóstico que se estaría complementando en la EED. Estos dos ASM se calificaron como cumplidos; no hay información de los otros ASM programados para 2012.

En 2013-2014 la Ficha de Evaluación realizada no informa sobre avances de los ASM. Sólo se definieron tres nuevos ASM específicos, relativos al programa Estímulos para el Desarrollo Humano y las Capacidades de los Adultos (EDHUCA) y uno institucional referido a la cobertura integrada de los programas alimentarios de Sedesol en los municipios de la Cruzada Nacional Contra el Hambre (CNCH).

En resumen, la información de Oportunidades disponible de 2007-2009 permite visualizar que sólo un pequeño número de recomendaciones se traduce en ASM (cuatro de 14 y 27 de 220). Las evaluaciones 2007-2010 tienden a insistir con recomendaciones dirigidas a la educación y a la salud, pero lo más sorprendente es que dos de las recomendaciones sustantivas más importantes de 2007-2008 sigan apareciendo como ASM pendientes en 2012-2013: la actualización del diagnóstico del programa y la elaboración de su programa

¹³ Los ASM fueron: actualizar el diagnóstico del programa, establecer el procedimiento para elaborar su programa estratégico, elaborar un agenda de evaluación externa de mediano plazo, identificar las potenciales sinergias con otros programas sociales federales, implementar un estrategia de capacitación y comunicación a los beneficiarios para mejorar la alimentación y la nutrición, mejorar los manuales de procedimiento del proceso de solicitud-entrega de apoyos, poner mayor atención en el análisis del logro académico de los becarios en las pruebas estandarizadas, y también en las acciones para atender a la población con base en la microzona.

estratégico. También se observa una importante reiteración de ASM sobre el tema de la definición de las poblaciones potencial y objetivo, la insuficiencia de recursos y la necesidad de aumentar la coordinación interinstitucional para desarrollar mayores sinergias.¹⁴ Dicha reiteración, junto con el limitado seguimiento numérico disponible, sugiere la idea de que los ASM de Oportunidades se están cumpliendo muy poco, pero esto no puede verificarse ya que los programas no aparecen desglosados en el informe de seguimiento de ASM que prepara anualmente el Coneval. Sin embargo, este problema no ha menoscabado el prestigio internacional de Oportunidades que, como se mencionó al inicio del capítulo, acaba de ser elegido como una de las nueve experiencias mundiales de evaluación dignas de relatarse por su contribución a la mejora de las condiciones de vida de sus beneficiarios en el concurso “Evaluations that make a difference”(EvalPartners, 2013).

En síntesis, la Sedesol tiene el mayor número de ASM institucionales y su nivel de cumplimiento se ha ido reduciendo notoriamente; su principal programa también muestra indicios en ese sentido.

- Educación¹⁵

En términos generales, en los últimos tres ciclos presupuestales, la SEP presentó el mayor número de ASM por dependencia o entidad con un total de 349, 97% de tipo específico, lo que facilita su cumplimiento porque no se necesita acordarlos con otras instancias (Cuadro 46). Los 38 ASM del último ciclo se originaron a partir de 45 recomendaciones de la evaluación correspondiente (84%), lo que lo ubica en el segundo lugar después de CDI (Cuadro 34); su cumplimiento puede ser informado hasta marzo de 2015. El volumen de ASM terminados ha ido aumentando: 40, 42 y 57% respectivamente, pero se recuerda que en este último ciclo la SEP fue la dependencia con más cancelaciones y bajas de ASM: 35 y 64 respectivamente (cuadros 39 y 40).

¹⁴ Esta última, por ejemplo, la podemos encontrar como recomendación de la evaluación de 2008 y 2010, y como ASM de 2008-2009, 2009-2010 y 2010-2011.

¹⁵ El texto correspondiente a este sector específico fue elaborado conjuntamente con la doctora Talina Olvera Mejía.

CUADRO 46
Número de ASM por tipo en la SEP

Tipo de ASM	Ciclo 2011-2012	Ciclo 2012-2013	Ciclo 2013-2014	Total de los tres ciclos
Específicos	127	176	36	339
Institucionales	5	3	2	10
Total	132	179	38	349

Fuente: elaboración propia con información del Coneval (2012e, 2013j y 2014b).

En cuanto a las cuatro metaevaluaciones realizadas en la segunda parte de este libro, una, Escuelas de Excelencia para Abatir el Rezago Educativo, sólo cuenta con una evaluación de diagnóstico sin recomendaciones, y otra, Escuela Digna, sólo recibió una recomendación en la EED de 2013 (cuantificar la población objetivo y atendida), que la dependencia informó en su documento de opinión institucional que ya contaba con una metodología para cumplirla. En consecuencia sólo se revisa el caso de las dos restantes.

El programa Escuelas de Tiempo Completo tiene el antecedente de haber sido uno de los ganadores del Premio a los Programas Federales comprometidos con el Proceso de Evaluación (Cuadro 42) por el cumplimiento de las recomendaciones surgidas de su evaluación de diseño 2008. Otro tanto sucede con el Programa de Escuelas de Calidad en relación con la evaluación Complementaria de 2008 y la EED 2009.

Como consecuencia de la EED de Escuelas de Tiempo Completo 2009-2010 se habrían generado tres ASM que recogen las recomendaciones formuladas en relación con la reelaboración de la MIR, la aplicación de instrumentos que permitan conocer el nivel de satisfacción de los beneficiarios y la llegada en tiempo y forma de los recursos y su adecuada focalización. El informe de seguimiento de ASM de 2011, absolutamente repetitivo, señala que ya se modificó la MIR y que se renueva periódicamente, que se instrumentó la aplicación de una Estrategia de Seguimiento Cualitativo de una muestra de escuelas, que no refiere la recolección de información de los beneficiarios, y que seguirá realizando informes trimestrales sobre el uso de los recursos financieros, desglosados por entidad, por medio del Sistema de Monitoreo y Acompañamiento de las Escuelas de Tiempo Completo (Simaetic), que tampoco cubre la solicitud del

ASM. En resumen, las dos últimas acciones no muestran cumplimiento alguno y la primera no especifica si los cambios introducidos a la MIR son los sugeridos por el evaluador (Coneval, 2011e); sin embargo, se consideró que los tres ASM estaban cumplidos.

La ECR de 2011 precisa que se aplicó una encuesta de opinión entre beneficiarios, las modificaciones realizadas a la MIR y la utilidad del Simaetic, indicando retrasos en el cumplimiento de los ASM (Coneval 2012d). Señaló también que el programa utilizaba regularmente los informes de evaluaciones externas para la toma de decisiones sobre su diseño y operación, y aprovechaba los resultados de esas evaluaciones para definir acciones que contribuyeran a mejorar su gestión y resultados, además de que había diseñado e implementado diversos mecanismos de seguimiento y evaluación en prácticamente todos los ámbitos de operación del programa.

En 2012, de las cuatro recomendaciones pendientes dos se habían atendido, una estaba en proceso y otra había sido rechazada argumentando su improcedencia por no ser atribución del programa (estudio de diagnóstico sobre las necesidades de alimentación). Respecto al avance de los ASM, los documentos señalaban un cien por ciento realizado de acuerdo con las fechas de término.

En 2014 el programa seguía sin tener una evaluación de impacto que fue programada en el PAE de 2012 y ha sido postergada para 2016. Su realización ha formado parte de las recomendaciones de diversas evaluaciones y se llegó a incluir como ASM, pero la dependencia argumenta no tener competencia para decidir el tipo de evaluación que realizará (lo que no es cierto en este caso en que debe proponerla al Coneval) y no tener recursos presupuestales para realizarla. Otras recomendaciones han sido rechazadas con argumentos similares: que sean claras, puntuales, factibles e indiquen los beneficios que implicaría su realización.

Además de los ASM específicos e institucionales, el programa ha sido objeto de un ASM interinstitucional en cada ciclo; el mismo se repite en los dos últimos ciclos, lo que evidencia que no ha sido cumplido.

La evaluación revisada de impacto cualitativo de Escuelas de Calidad 2012-2013 se hizo al margen del sistema organizado por el Coneval, por lo que no contó con ASM, pero tampoco aportó resultados relevantes. Sin embargo, la evaluación específica realizada en 2013 dio lugar a ocho ASM, dos de los cuales se vinculan con la focalización y el mejoramiento de la MIR, que ya estaban planteados desde 2011 (Coneval, 2011f). Los demás incluyen transparencia

en la aplicación de recursos, criterios de gestión homogéneos por entidades federativas, simplificación de la carga administrativas en centros escolares y realización de una evaluación de procesos (SEP, 2014b). Ninguno de estos ASM se dirige a asuntos sustantivos del programa, por lo que más que una evaluación de procesos se requiere una efectiva evaluación de impacto. Adicionalmente, el programa ha sido objeto de un mismo ASM interinstitucional en cada uno de los dos últimos ciclos (justificación incompleta de todas sus metas), lo que muestra que no se había cumplido.

En síntesis, la SEP tiene un alto porcentaje de recomendaciones transformadas en ASM en 2013-2014, pero también concentra el mayor número de éstos y un cumplimiento de sólo 57% de los mismos, a pesar de haber concentrado 46% del total de ASM cancelados y dados de baja.

- Medio ambiente¹⁶

En los últimos tres ciclos presupuestales, el conjunto de programas de la Semarnat que fueron evaluados se comprometieron a realizar 181 ASM, ocupando el tercer lugar en el conjunto de dependencias y entidades (Cuadro 47). Los 39 ASM del último ciclo provienen de 103 recomendaciones recibidas, lo que significa que apenas más de la tercera parte (38%) fue efectivamente considerada para mejorar los programas, ocupando el tercer lugar después de la Sedesol y Economía (Cuadro 34). El volumen de ASM terminados ha venido fluctuando: 36, 76 y 47% respectivamente. Se recuerda que en este último ciclo la Semarnat fue la dependencia con mayor número de bajas de ASM después de la SEP: 57 (Cuadro 40).

En cuanto a las tres evaluaciones analizadas en la segunda parte del libro, la relativa al PNANP fue ajena al sistema del Coneval, por lo que ese caso no dio lugar a ASM. En materia de ANP, dicho Consejo sólo evalúa regularmente tres programas: vigilancia comunitaria en ANP y zonas de influencia desde 2011, Promovi y Promanp desde 2012. Como en la citada evaluación sólo se rediseñó el sistema de indicadores y se desconoce que se haya hecho una nueva al término del programa en 2012 como se había planeado, o bien que las propias ANP

¹⁶ El texto correspondiente a este sector específico es una contribución de las doctoras Laura Rodríguez Cardozo y Esthela Sotelo Núñez.

CUADRO 47
Número de ASM por tipo en la Semarnat

Tipo de ASM	Ciclo 2011-2012	Ciclo 2012-2013	Ciclo 2013-2014	Total de los tres ciclos
Específicos	63	44	28	135
Institucionales	5	30	11	46
Total	68	74	39	181

Fuente: elaboración propia con información del Coneval (2012e, 2013j y 2014b).

estén siendo evaluadas. Sólo resta señalar que el SIMEC ha informado con base en el conjunto de nuevos indicadores en 2010-2013 a excepción del relativo al porcentaje del territorio nacional protegido; en cambio ha agregado tres indicadores más: ANP que implementan alguno de los elementos del Programa Nacional de Cultura para la Conservación, incremento en la participación de la sociedad en acciones de conservación y efectividad de la integración de sistemas de ANP estatales, municipales y otras modalidades de conservación en coordinación con la Conanp (SIMEC, 2013).

En relación con el Promanp, cuya evaluación de diseño también fue analizada, se mencionó que se generaron cinco recomendaciones respecto de la estructura y gestión del programa (población potencial y objetivo, incongruencias en la MIR, insuficiente información, etcétera) que, si bien fueron aceptadas en la posición institucional, sólo tres de ellas se transformaron en ASM y se reportaron cumplidas (Semarnat, 2013). Sin embargo, al año siguiente su evaluación específica muestra dos recomendaciones: la primera repite parte de las señaladas por la evaluación de diseño al proponer una definición y cuantificación de la población potencial y objetivo, lo que indica que se mantenía su ambigüedad, y la segunda sugiere revisar la factibilidad de que el programa cuente con otras evaluaciones externas en sus diferentes modalidades. Frente a éstas la respuesta institucional fue que se propondría una definición y se cuantificarían las poblaciones referidas y que se harían gestiones para contar con recursos adicionales en el año 2014 de manera de poder contratar una evaluación externa. Todavía no hay información publicada sobre su cumplimiento. Se debe agregar que el programa podría cumplir plenamente su cometido en 2015 ya que el propio titular de la Semarnat ha informado que a fines de ese año se contaría

con todos los Programas de Manejo y Conservación (PMYC), cuya realización constituye la razón de ser del Promanp (Semarnat en Noticias MVS, 2014).

Por otra parte, de la evaluación del diseño del programa “Estabilización de cuencas y acuíferos” se derivaron una serie de recomendaciones puntuales. De éstas, sólo la recomendación de “Replantear el objetivo de Fin de la MIR” fue considerada como un ASM. En el seguimiento correspondiente (Coneval, 2014b) se sugiere revisar el objetivo de la MIR, considerando que el programa cambió su nombre por el de “Rehabilitación y modernización de infraestructura de riego y temporal tecnificado”. Éste ASM se declaró terminado el 30 de abril de 2014.

Además, algunas de las recomendaciones propuestas se rescatan en el documento “Diagnóstico del programa presupuestario K141. Rehabilitación y Modernización de la infraestructura de riego y temporal tecnificado”, realizado por la misma Conagua y publicado por Coneval en 2014.¹⁷ Este documento diagnóstico menciona que el objetivo del ejercicio es justificar la permanencia, ampliación o modificación sustantiva del programa, pero no aclara cuál de las tres posibilidades se propone alcanzar. Sin embargo, desde el cambio de nombre que sufrió el programa, seguido por las estrategias de implementación, se conoce que la evaluación de diseño contribuyó a una modificación sustantiva.

El documento diagnóstico señala que este cambio persigue, de entrada, la disposición de un mayor monto destinado a gastos de operación, enfocados al mantenimiento de la infraestructura de riego. El argumento principal que se ofrece es que, de acuerdo con los resultados arrojados por la evaluación del diseño, el nombre del programa no correspondía a las acciones que realmente se llevan a cabo, que no tienen que ver con la conservación de los acuíferos, sino con rehabilitar, modernizar y conservar la infraestructura hidroagrícola.

Se señala que, independientemente del cambio de nombre, las acciones que se llevan a cabo dentro del programa siguen siendo las mismas de antes, y que, de no realizarse, el recurso hídrico no se otorgaría con oportunidad a las áreas de riego. Sin embargo, se considera que el cambio de nombre no es trivial, y trae consigo implicaciones de distinta índole. En su diseño original, el nombre del programa respondía a una necesidad de conservación ecológica, no productiva. En este tenor, las recomendaciones emitidas por el evaluador en la MIR buscaban específicamente que los indicadores se alinearan más a este

¹⁷ [http://www.coneval.gob.mx/Informes/Evaluacion/Diagnostico/Diagnostico_2014/Diagnostico_2014_SEMARNAT_K141.pdf], fecha de consulta: 3 de diciembre de 2014.

principio de conservación de recursos. En el documento de diagnóstico que modifica al programa, se incorporaron al pie de la letra las recomendaciones hechas por el evaluador, pero ahora bajo una etiqueta que abiertamente señala que no persigue un fin necesariamente ecológico, sino de productividad. En otras palabras, antes se tenía un programa con un nombre que denotaba un fin de conservación ecológica, pero cuyos indicadores eran totalmente enfocados a fines productivos, y ahora se tiene un programa con fines abiertamente productivos y de conservación de infraestructura, con indicadores que mezclan los temas de productividad con los de conservación ecológica.

Por último, es posible afirmar que la evaluación de diseño sí brindó información para modificar el programa, pero que la orientación o lectura que se dio a esta información fue a discreción de las áreas responsables. Se considera que responder con un cambio de nombre a las recomendaciones emitidas por la evaluación de diseño realizada no justifica que los resultados de la evaluación hayan sido utilizados. En última instancia, ni siquiera justifica que se haya emprendido todo un ejercicio evaluativo para llegar a la conclusión de que había que cambiar “a modo” el nombre del programa.

En síntesis, la Semarnat retoma pocas recomendaciones en sus ASM y de éstos cumple con un porcentaje reducido. La revisión de los tres casos considerados muestra una utilización relativa: se utilizan los indicadores diseñados para informar sobre las ANP pero éstas siguen sin evaluarse al igual que el programa que las normaba, se aceptan las recomendaciones del Promanp pero los evaluadores consideran que no se han resuelto, y otro tanto sucede con el cambio de nombre del programa de Estabilización de Cuencas y Acuíferos, que genera una nueva incongruencia.

- Salud

El conjunto de programas de la SS que fueron evaluados se comprometieron a realizar un número pequeño de ASM (41) en los últimos tres ciclos presupuestales (Cuadro 48). Los 18 ASM del último ciclo provienen de 39 recomendaciones recibidas, lo que constituye 46% (Cuadro 34). El porcentaje de ASM terminados se ha ido reduciendo: 71, 60 y 44 respectivamente. La SS sólo canceló dos ASM específicos de 2013-2014 y no dio de baja a ninguno (Cuadro 40).

CUADRO 48
Número de ASM por tipo en la SS

Tipo de ASM	Ciclo 2011-2012	Ciclo 2012-2013	Ciclo 2013-2014	Total de los tres ciclos
Específicos	5	19	14	38
Institucionales	0	1	2	3
Total	5	20	16	41

Fuente: elaboración propia con información del Coneval (2012e, 2013j y 2014b).

En cuanto a las dos evaluaciones analizadas en la segunda parte del libro, la relativa al Seguro Popular de Salud 2007 sólo había efectuado una recomendación relativa al establecimiento de un modelo integral de monitoreo y evaluación, si bien también había identificado problemas en la construcción de la MIR y, en especial, había señalado falta de seguimiento de las recomendaciones de los dos años previos a la evaluación. En 2008 se le aplicó una evaluación de impacto que arrojó resultados globales positivos y señaló algunas oportunidades de mejora que incluían cambios en procedimientos de operación identificados como débiles en otras evaluaciones. Éstos podrían recoger la recomendación de 2007, que reaparece como compromiso del programa en la EED 2009-2010. La siguiente EED no retoma el tema pero vuelve a incluir como una de sus recomendaciones: definir e instrumentar indicadores de seguimiento del acceso a servicios y su impacto en la salud. En 2012 se señala que los anteriores han sido resueltos y en 2013 la debilidad identificada recae en falta de indicadores de utilización de los servicios. En síntesis, la reaparición de asuntos recomendados no permite tener claridad sobre su pleno cumplimiento.

Sicalidad, el otro programa revisado, fue en 2011 el ganador del Premio a los Programas Federales comprometidos con el Proceso de Evaluación (Cuadro 42), dirigido a los funcionarios que utilizaron las evaluaciones en los programas federales que dirigen. Para ello se tomaron en cuenta las evaluaciones de diseño 2008, EED de 2009 y 2010 y procesos de 2010. Entre los 10 ASM provenientes de las anteriores que justificaron su pleno cumplimiento, Montiel (2012) destaca: *a*) involucrar a los pacientes en la supervisión de la calidad de los servicios médicos; *b*) medir la cultura de la calidad en las unidades médicas participantes, así como el impacto del programa; *c*) acreditar a las unidades

médicas en temas de calidad de los servicios médicos y *d*) ordenar los procesos de gestión y difusión del programa.

Sin embargo, no es tan claro su cumplimiento posterior a 2010. En el caso de la evaluación 2008-2009 las tres principales recomendaciones se dirigieron a la gestión de recursos humanos, la comunicación de razones de aceptación o rechazo de propuestas y el rediseño de la página web del programa. La convocatoria de 2010 incluyó el fortalecimiento de la figura del responsable estatal de calidad, que se había sugerido dentro de la primera recomendación presentada. La EED de ese mismo año sugirió nuevos temas: revisión de la MIR, establecer un mecanismo de evaluación de efectos o resultados e institucionalizar las líneas de acción más consolidadas antes de aspirar a desarrollar. El programa aceptó las dos primeras para 2011 y declaró inaplicable la tercera.

En 2013-2014, Sicalidad fue objeto de un ASM interinstitucional relacionado con la adopción, difusión y capacitación en las guías del Sistema Integral de Calidad, de cuyo cumplimiento aún no se ha publicado información.

En resumen, la SS retoma un porcentaje bajo de las recomendaciones que recibe y tiene un nivel de cumplimiento medio que muestra tendencia a disminuir.

- Desarrollo rural¹⁸

En términos generales, en los últimos tres ciclos presupuestales los ASM acordados por la Sagarpa son poco numerosos (Cuadro 49) y le permiten lograr un cumplimiento pleno y no recurrir a cancelaciones ni bajas de ASM. En cambio, los 17 del último ciclo se originaron a partir de 29 recomendaciones de la evaluación correspondiente, 59% de ésta (Cuadro 34). La tendencia ha sido pasar de un énfasis en ASM institucionales a su total desaparición, lo que facilita su cumplimiento debido a que para los ASM específicos sólo es necesaria la participación de la DGPE de la Sagarpa en conjunto con los encargados de sus programas (las unidades responsables). La mayoría de las recomendaciones que llegan a ser consideradas para ser un ASM tienen que ver con aspectos vinculados a mejoras a las reglas de operación, documentos normativos o en

¹⁸ El texto correspondiente a este sector específico es una contribución del maestro Osmar Medina Urzúa.

CUADRO 49
Número de ASM por tipo en la Sagarpa

Tipo de ASM	Ciclo 2011-2012	Ciclo 2012-2013	Ciclo 2013-2014	Total de los tres ciclos
Específicos	12	0	17	29
Institucionales	10	0	0	10
Total	22	0	17	39

Fuente: elaboración propia con información del Coneval (2012e, 2013j y 2014b).

el diseño en general del programa; de ahí también que se observe un amplio porcentaje de ASM específicos.

Sorprende que el número total de ASM del ciclo 2013-2014 no disminuyó significativamente, como sucedió en SEP y Semarnat, a pesar de que, como ya se mencionó, en el ejercicio fiscal 2014 se efectuaron cambios sustantivos en la estructura programática de la Administración Pública Federal, lo que provocó la salida de programas del PEF y por tanto, que no estuvieron obligados a realizar evaluaciones externas ni a incorporar nuevos aspectos susceptibles de mejora.

Conforme a las conclusiones y recomendaciones presentadas concretamente en las evaluaciones metaevaluadas, de la ECR 2011-2012 del PDCITYER se derivaron dos ASM: a) instrumentar la medición del grado de satisfacción de los beneficiarios y b) realizar un taller de MML para identificar y definir el problema central que el programa pretende atender. Del mismo modo, la ECR 2011-2012 de Procampo dio lugar a un ASM idéntico al primero de los dos anteriores.

Los ASM fueron cumplidos por la Sagarpa mediante una evaluación de satisfacción de los beneficiarios del programa y al mejorarse la MIR con base en modificaciones hechas al Árbol de Problemas. Sin embargo, la Dirección General Adjunta de Planeación y Evaluación de la Sagarpa considera que las conclusiones y recomendaciones derivadas de la ECR son muy generales y que no realizan hallazgos que no se pudiesen haber identificado de otra forma. Además de que, dada la estructura de los programas y las necesidades y expectativas de información identificadas, la aplicación del instrumento se realiza para cumplir

con la normatividad establecida pero se requiere obtener información adicional de utilidad inmediata para los operadores y encargados de los programas.¹⁹

Los términos de referencia de las ECR 2011-2012 establecen como responsabilidad y compromiso de los evaluadores externos que, de serles solicitado, acompañen a la dependencia en la presentación de los resultados de la evaluación. Pero, de los dos evaluadores externos que fueron entrevistados, ninguno tuvo participación como expositor de los resultados de la evaluación ante otra entidad que no fuese la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación.

Respecto al uso de la evaluación por parte de otros actores, todas las evaluaciones sirven al Coneval para hacer comparaciones entre los programas de una o varias dependencias, o hacer comparaciones retrospectivas del mismo. También se encontró que el Consejo Mexicano para el Desarrollo Rural Sustentable de la Cámara de Diputados revisa los resultados de las evaluaciones²⁰ pero no hubo evidencia de que se utilizaron como insumos para la toma de decisiones en la asignación del presupuesto o en el diseño o rediseño de los Programas y la Política de Desarrollo Rural y Alimentaria. Se considera que uno de los principales factores que ha limitado el uso de las evaluaciones es que no se ha cumplido con las expectativas y necesidades de información de los tomadores de decisiones el programa.

En conclusión, en el caso de la Sagarpa, el uso de las evaluaciones es limitado al momento de transformar las recomendaciones en ASM, después los acordados se cumplen plenamente.

A nivel del Gobierno del Distrito Federal

La utilización en el gobierno de la capital sigue la misma distribución que la federal en cuanto a su importancia: la mayor es de tipo instrumental, la siguiente de orden presupuestal, y la conceptual presenta similares problemas de medición.

¹⁹ Información derivada de las entrevistas aplicadas.

²⁰ La aseveración se hace con base en que se encontraron referencias a ellas en diversos calendarios de sesiones, actas, acuerdos y circulares en el Consejo Mexicano para el Desarrollo Rural Sustentable de la Cámara de Diputados [<http://www.cmdrs.gob.mx/actas/Paginas/default.aspx>], fecha de consulta: 4 de mayo de 2014.

- Utilización instrumental

Respecto de la utilización instrumental, el número de recomendaciones enviadas por Evalúa DF fue siempre mayor que las sugeridas en los informes de evaluación, mientras que en Coneval, como ya fue analizado, sólo una pequeña parte de las recomendaciones se convierte en ASM, impactando lógicamente en su nivel de aceptación y cumplimiento. Otra diferencia importante ya señalada es que las recomendaciones del Evalúa DF tienen carácter vinculatorio; sin embargo, las dependencias evaluadas tienen la posibilidad de manifestar su desacuerdo, en cuyo caso Evalúa DF puede controvertirlas ante la Comisión Intersecretarial de Desarrollo Social.²¹

En general, el cumplimiento total o parcial de las recomendaciones de 2009 a las evaluaciones iniciadas en 2008 fue bastante elevado hasta 2011,²² aunque existía un importante grado de retraso en su implementación y la entrega de información. Era posible afirmar que la aceptación de las recomendaciones ligadas a la operación de los programas o implementación de las políticas era mayor que la ligada a su reformulación. Las primeras incluyen fundamentalmente cambios en los procedimientos en curso, mientras que las últimas modifican la definición misma de la política y eso, en muchos casos, implica mayor disponibilidad de recursos. El Cuadro 50 muestra la evolución de las recomendaciones 2009-2011 y el siguiente el nivel de cumplimiento de las recomendaciones 2009 en junio de 2011.

En 2009 las 124 recomendaciones fueron dirigidas a 12 programas y dependencias, se distribuyeron aproximadamente en la misma proporción en materia 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 éstas implicaban una fuerte repercusión presupuestal. Todas las controvertidas correspondían a dos de los programas evaluados (Evalúa DF, 2011g).

²¹ La primera versión de este análisis se publicó en Cardozo (2013).

²² Posteriormente no se ha publicado información, razón por la que no ha sido posible cubrir hasta el 2014, como se hizo a nivel federal. La información fue solicitada por internet pero no se recibió respuesta.

CUADRO 50
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: Evalúa DF, Primer informe semestral del estado de las recomendaciones aceptadas, junio de 2011.

CUADRO 51
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: Evalúa DF, Primer informe semestral del estado de las recomendaciones aceptadas, junio de 2011.

De las correspondientes al programa emblemático para Adultos Mayores fueron rechazadas siete de 19. Cinco se vinculaban con la propuesta de garantizar el derecho al acceso gratuito a los servicios médicos y medicamentos a las personas residentes en el Distrito Federal que carecen de seguridad social laboral, a través de los servicios públicos de salud del GDF. Una sexta corresponde a la propuesta de reducción de causales de baja del programa y la última, a la revisión de los propósitos del procedimiento de visitas domiciliarias.

Por otra parte, en las recomendaciones aceptadas en 2009, al 30 de junio de 2011 (último informe disponible) había un cumplimiento global de aproximadamente 73%, suponiendo que el cumplimiento parcial fuera del 50%. El Programa de Pensión Alimentaria de Adultos Mayores, de las 12 recomendaciones que había aceptado, contaba con ocho cumplidas, dos parcialmente cumplidas, y dos incumplidas. A título de ejemplo y por tratarse del

principal programa del GDF, al igual que se hizo con Oportunidades a nivel federal, éstas se desglosaban, sintetizando su contenido, en:

CUMPLIDAS

- Formular una nueva versión de la Carta Compromiso en la cual se eliminen las cláusulas ambiguas y que tome en cuenta las modificaciones que se recomienda introducir en el Reglamento.
- Modificar el estatuto laboral de capacitadores y coordinadores cambiando su contratación por honorarios asimilados a salarios y por la de plaza de estructura; y aumentar la remuneración de los capacitadores.
- Realizar un diagnóstico de funciones y puestos de la estructura central de la Dirección de Operación de la Atención a los Adultos Mayores (DOAAM) y gestionar los recursos que resulten necesarios en función del mismo.
- Gestionar la adquisición o transferencia de equipo de cómputo adecuado y suficiente para la operación tanto central como regional de la pensión alimentaria.
- Suscribir un convenio con el Registro Civil del Distrito Federal para que proporcione regular y oportunamente al Instituto para la Atención de Adultos Mayores (IAAM-DF) copia de las actas de defunción correspondientes a la población cubierta por la pensión.
- Ampliar la información estadística sobre la pensión alimentaria en el portal de internet del IAAM-DF.
- Incorporar a dicho portal un listado de nombres y direcciones de los Centros de Salud donde se localizan los módulos de atención, incluyendo (en su caso) el teléfono del módulo respectivo y su dirección electrónica.
- Reformular la “Encuesta de percepción de los adultos mayores con relación a la pensión alimentaria”, con el objeto de incrementar su utilidad como instrumento para la evaluación del diseño, la operación y el impacto social de la pensión alimentaria.

PARCIALMENTE CUMPLIDAS

- Mejorar las condiciones de la operación regional en lo que respecta a los módulos (y su equipamiento) de atención a adultos mayores localizados en Centros de Salud del Distrito Federal.
- Desarrollar un sistema permanente de evaluación de la operación con base en la mayor información disponible en las bases de datos en las que se sustenta el registro y seguimiento de la operación, incluido el padrón de beneficiarios.

INCUMPLIDAS

- Incorporar en el artículo 17 del Reglamento una cláusula que especifique un plazo máximo de 10 días para la entrega de la tarjeta a los derechohabientes dados de alta en el padrón de beneficiarios si acuden al módulo a recogerla, y de 20 días para los que requieran, por su falta de autonomía, que les sea entregada en su domicilio.
- Formular el Programa de Atención Integral al Adulto Mayor con una perspectiva temporal de 2011-2012.

En 2010 las tres recomendaciones no aceptadas correspondieron al Instituto de las Mujeres del Distrito Federal; 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 programa se hizo la primera experiencia de dirigir recomendaciones directamente a la Secretaría de Finanzas cuando existían importantes repercusiones en las necesidades financieras. Se trató concretamente de la recomendación relacionada con el traslado de recursos de este programa a otro de nueva creación, lo que fue aceptado; sin embargo, no se trataba de una ampliación de recursos, sino de una mera transferencia.

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 trató de siete grandes recomendaciones, desglosadas en 21 para su seguimiento en el caso de agua potable (Evalúa DF,

2010h) y de 10, subdivididas en 56, en materia de empleo (Evalúa DF, 2010i). Las primeras fueron dirigidas a la Sedema, al Sistema de Aguas de la Ciudad de México (SACM), a la Secretaría de Desarrollo Urbano y Vivienda (Seduvi), a la Secretaría de Finanzas, a la Contraloría General del Distrito Federal y a la Jefatura de Gobierno; mientras que las últimas lo hicieron a la Secretaría de Trabajo y Fomento al Empleo, a la Secretaría de Desarrollo Económico, a la Secretaría de Finanzas, a la Secretaría de Obras y Servicios, a la Oficialía Mayor y a la Jefatura de Gobierno (Evalúa DF, 2011l). Fueron rechazadas 8 de las 21 en el caso de agua potable.

Como puede verse, en 2011 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 ésta empezó también a rechazarlas por considerar que no le competía emitir opinión sobre las recomendaciones de carácter operativo, aunque éstas tuvieran fuerte impacto presupuestal.

Adicionalmente a la información presentada, se recuerda que las tres meta-evaluaciones realizadas a las políticas evaluadas en 2010-2011 en la segunda parte del libro arrojaron 47% de aceptación plena y 20% de aceptación parcial (Cuadro 30), con lo que se estima 57% de aceptación global, bastante inferior a los porcentajes reportados en años previos (Cuadro 50).

En síntesis, el carácter vinculatorio del Evalúa DF no le ha permitido, hasta la fecha, tener la contundencia esperada en materia de reformulación de políticas y programas. En cambio, la experiencia de la Sedema, cuyo caso se revisa a continuación, fue exitosa con la particular característica de que las modificaciones al programa fueron encargadas a sus mismos evaluadores.

Pero, además de asegurar la puesta en práctica de las buenas recomendaciones surgidas de un proceso de evaluación, es necesario verificar el impacto que las mismas 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.

- Un caso especial²³

Además de las evaluaciones coordinadas por Evalúa DF en la segunda parte del libro se incluyó la realizada al PACCM 2008-2012 a solicitud de la Sedema. A continuación se revisa brevemente el nivel de utilización de esta última.

La evaluación mencionada fue entregada en noviembre de 2012, posteriormente la Dirección del Programa de Cambio Climático y Proyectos MDL, como colaboradores de la Secretaría Técnica de la CICCDF, invitó al CMM a informar y aclarar dudas de los resultados de los estudios realizados a los integrantes de dicha Comisión.²⁴

En junio de 2014 el GDF dio a conocer los dos instrumentos de planeación que orientarán la política de cambio climático durante el gobierno de Miguel Ángel Mancera, por un lado, la Estrategia Local de Acción Climática de la Ciudad de México 2014-2020 (ELACCM) y por el otro, el Programa de Acción Climática de la Ciudad de México 2014-2020. Ambos documentos fueron elaborados por el propio Centro (CMM, 2014 y Sedema/CMM, 2014) contando con la colaboración de los dos principales responsables de la evaluación analizada.

Desde la ELACCM se retomaron algunas de las recomendaciones que el propio CCM emitió a partir de la evaluación realizada al PACCM 2008-2012. Destaca la que se refiere a contar con una línea base de mitigación, adaptación, educación y comunicación a partir de un inventario de emisiones de gases de efecto invernadero consolidado para el Distrito Federal, y que ésta se desarrolle conforme a las directrices del IPCC (CMM, 2012), cuya realización se efectuó en los términos señalados, con base en la versión 2006 del IPCC y en las Guías de Buenas Prácticas del IPCC 2000 (Sedema/CMM, 2014). Otra de las recomendaciones que se retomó es la de elaborar un análisis de vulnerabilidad o diagnóstico que oriente y justifique la instrumentación de las acciones de adaptación.

En cuanto al Programa de Acción Climática de la Ciudad de México 2014-2020 (Sedema/CMM, 2014), su diseño se apegó a las observaciones y

²³ El texto correspondiente a este caso es una contribución de la doctora Angélica Rosas Huerta.

²⁴ Información proporcionada por el encargado de la Oficina de Información Pública de la SMA del GDF a partir de la solicitud número de folio 112000162514.

recomendaciones emitidas en la evaluación al PACCM 2008-2012.²⁵ Entre las sugerencias retomadas se encuentran las siguientes: el establecimiento de la línea base de mitigación, adaptación, educación y comunicación antes mencionado y el diseño de una herramienta informática que permita dar seguimiento a las acciones contempladas en el programa, así como evaluarlas. Todas las acciones tienen asignado un alcance, una meta, responsables y colaboradores así como indicadores de seguimiento e impacto.²⁶ También se establece una relación explícita y directa entre las acciones orientadas a la mitigación y las categorías clave del inventario de emisiones; así como entre las acciones orientadas a la adaptación y las principales amenazas físicas y factores sociales de vulnerabilidad ante el cambio climático.

Destaca también la priorización que se realizó de las acciones contenidas en el PACCM 2014-2020, como otra de las recomendaciones realizadas por el grupo evaluador. De igual forma, la depuración de acciones que se realizó para seleccionar aquellas que se incluyeron en el programa tomando en cuenta las consideraciones mostradas en la evaluación del programa.

La evaluación del PACCM 2008-2012 contribuyó en la mejora del diseño del Programa de Acción Climática de la Ciudad de México 2014-2020,²⁷ componente importante de la política en este campo, pero no el único.

- Utilización presupuestal

En el caso del GDF, si bien no se cuenta con la información detallada por programa, menos aún cuando se trata de políticas, ésta existe en algunos casos

²⁵ Información que se recupera a partir de la revisión del PACCM 2014-2020 y ratificada por el funcionario responsable de la Oficina de Información Pública de la SMA del GDF a partir de la solicitud con número de folio 112000162514.

²⁶ Estos indicadores permitirán realizar una evaluación del programa a partir de los sistemas MRV (Monitoreo, Reporte y Verificación), metodología que también fue utilizada en el estudio de trazabilidad de la información que realizó el grupo evaluador.

²⁷ Se debe destacar que es un caso *sui generis* puesto que la planeación no es realizada por el gobierno en función de la evaluación sino encargada al mismo centro de investigación.

en que el programa y la institución que lo opera prácticamente se identifican, lo que nos permite acercarnos a su análisis.

El Cuadro 52 muestra la evolución del presupuesto del GDF por organismos públicos en 2008-2012.²⁸ Como podemos ver, en precios corrientes, el incremento promedio fue de 17%, pero como la inflación acumulada del periodo fue aproximada a 21%, casi se mantuvo estable en términos reales.

CUADRO 52
*Gasto social ejercido en los principales organismos públicos del GDF
2008-2012 (millones de pesos corrientes)*

Organismos públicos	2008	2009	2010	2011	2012	% incremento 2008-2012
Instituto Atención de Adultos Mayores	3 945.5	4 220.4	4 826.2	5 431.5	5 477.2	0.39
Servicios de Salud	3 030.3	3 119.4	3 048.7	3 915.8	3 319.5	0.09
Desarrollo Integral de la Familia	2 777.3	2 375.9	2 621.8	2 745.6	2 724.8	-0.02
Instituto de Vivienda	2 409.8	2 204.3	2 612.4	2 635.4	2 588.1	0.07
Fideicomiso de Educación Garantizada	1 309.0	1 643.6	1 602.6	1 632.9	1 701.5	0.30
Instituto Nacional de las Mujeres	143.1	133.2	136.8	143.3	106.7	-0.25
Instituto de la Juventud	90.1	81.9	87.9	110.0	90.8	0.01
Total	13 705.1	13 778.7	14 936.4	16 614.5	16 008.6	0.17

Fuente: adaptación propia a partir de datos de la Dirección de Estadística del Evalúa DF, calculados con base en las Cuentas Públicas del GDF para 2008-2010, y con los decretos de presupuesto aprobados en 2011-2012.

²⁸ Evalúa DF no ha publicado los cálculos para años posteriores.

Aunque existen variaciones pequeñas entre organismos (nada comparable a las analizadas en el caso federal), identificamos dos instancias que aumentan un poco más que la media: el Fideicomiso para la Educación Garantizada, Fidegar (0.30%) y el IAAM (0.39%) (Evalúa DF, 2010g). Pero en estos casos existen explicaciones que nada tienen que ver con el éxito mostrado por la gestión de la política o programa en sus evaluaciones: el primero tuvo a su cargo el Programa del Bachillerato Universal, Prepa Sí (Evalúa DF, 2009b) y fue el único que no aceptó ninguna recomendación del Evalúa DF; de igual forma, en el segundo caso fueron rechazadas las recomendaciones de fuerte impacto presupuestal y el incremento debe estar relacionado más bien con la reducción de la edad para recibir el apoyo (de 70 a 68 años), que aumentó su población objetivo.

Evalúa DF debe entregar copia de los informes de evaluación a la Comisión de Desarrollo Social de la Asamblea Legislativa del DF y ésta los ha utilizado en diversos casos, más 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, el proceso de aceptación y cumplimiento de las recomendaciones no muestra impactos sustantivos para mejorar la distribución de los recursos presupuestales en ninguno de los dos casos considerados, federal ni del Gobierno del Distrito Federal.

- Utilización conceptual

A pesar de que todos los documentos relativos al proceso de evaluación han estado accesibles desde el inicio de los trabajos del Evalúa DF (TR, informes, recomendaciones, respuestas institucionales, controversias, etcétera) con la intención de generar un proceso de aprendizaje conceptual, no existen estudios que midan los avances logrados en ese sentido, por lo que la situación es muy similar a la descrita en materia federal.

Tal vez uno de los mejores ejemplos de este tipo de aprendizaje o utilización sea la concepción misma del Evalúa DF en 2008, diferenciada de la del Coneval del 2006 en múltiples aspectos normativos, teóricos, metodológicos y organizativos, que se resumen en los cuadros 53 y 54 (Cardozo, 2011).

CUADRO 53
Enfoques de la evaluación

Coneval: énfasis positivista	Evalúa DF: énfasis constructivista:
Busca descubrir una verdad única y objetiva con un método similar al de las ciencias físico-naturales.	Considera que existen múltiples y subjetivas verdades, identificables por métodos diversos en ciencias sociales.
Considera que todos los programas deben ser evaluados en forma similar.	Sostiene que cada evaluación debe construirse en función de las características específicas del programa.
Aspira a la realización de estudios experimentales como la forma más rigurosa de identificar los impactos de un programa.	Considera que los estudios experimentales son difíciles y costosos, no han dado los resultados técnicos esperados y tienen importantes connotaciones éticas negativas.
Valora la cuantificación (medición y pruebas) muy por encima de la aplicación de métodos cualitativos.	Valora la contribución tanto de los métodos cuantitativos como cualitativos.
Promueve la realización de trabajos por expertos, en detrimento de la participación social.	Realiza trabajos especializados pero promueve la participación social para generar aprendizaje ciudadano y credibilidad de la evaluación.
Prioriza el análisis disciplinario de corte económico.	Enfatiza la necesidad de contar con equipos interdisciplinarios pertinentes al programa.

Fuente: Cardozo (2011).

CUADRO 54
Prácticas de evaluación

Práctica del gobierno federal	Práctica del gobierno del Distrito Federal
Proceso implantado sin previa preparación.	Impartición de cursos, talleres, diplomados en evaluación y elaboración de lineamientos.
Evaluación anual de todos los programas gubernamentales sujetos a Reglas de Operación.	Evaluación anual de algunas políticas y programas seleccionados con base en criterios explícitos.
Evaluación anual de todos los aspectos del programa, según la Ley General de Desarrollo Social.	Evaluación de algunos aspectos de los programas seleccionados.
Convocatoria por invitación a por lo menos tres organizaciones.	Convocatoria por concurso abierto.
Posible acumulación de múltiples trabajos de evaluación por un mismo equipo.	Limitación del número de evaluaciones a realizar al mismo tiempo.
Pago efectuado con el presupuesto del propio programa evaluado.	Creación de un fondo independiente para el pago de las evaluaciones.
Recomendaciones no vinculatorias.	Recomendaciones vinculatorias.

Fuente: Cardozo (2011).

Conclusiones de la tercera parte

Myriam Cardozo Brum

El primer capítulo aportó un amplio marco de análisis teórico e investigaciones empíricas mundiales para comprender mejor los determinantes de la utilización de las evaluaciones que van desde el clima político, la cultura de evaluación o la credibilidad de los evaluadores hasta asuntos mucho más concretos como la presentación, discusión y canales de distribución del informe o la presencia de incentivos para adoptar las recomendaciones. La riqueza de propuestas es mucha pero el consenso sobre los factores más relevantes es muy limitado; gira sobre todo en torno a la cultura de evaluación y el involucramiento de los destinatarios del trabajo evaluativo durante su proceso. En cambio, pocos autores mencionan la calidad de informe y de la metodología aplicada.

Después de recoger brevemente las pocas experiencias internacionales exitosas revisadas, se pasa al análisis de la experiencia mexicana. A nivel federal y en relación con la utilización instrumental es donde se tiene mayor información disponible. El número de ASM por programa (calculado con los datos del Cuadro 35) muestra una clara tendencia a la disminución, con un aumento pasajero en 2011-2012, provocado por la realización de ECR. En consistencia con lo anterior, su nivel de cumplimiento ha ido en aumento hasta alcanzar 82%, a pesar del carácter no vinculatorio de las decisiones del Coneval, con una disminución en 2012-2013 que coincide con la aplicación de las recomendaciones surgidas de la ECR. Los ASM se concentran cada vez más en las correcciones de programas, mientras las reorientaciones vinculadas con el rediseño se mantienen

en sólo 22% (Cuadro 38). A su vez, el porcentaje de ASM específicos en el total aumentó considerablemente a mitad del periodo 2007-2013, pero ha vuelto a descender levemente, conservando 9% de aumento entre sus extremos (Cuadro 36). También han tendido a disminuir los ASM de alta prioridad (Cuadro 37).

La Sedesol y la Semarnat son las dependencias que recibieron mayor número de recomendaciones y, junto con la SEP, las que emitieron la mayor cantidad de ASM en 2012-2013, pero sus porcentajes de transformación de unas en otros sólo alcanzan 28 y 38% respectivamente (Cuadro 34). Sedesol cuenta también con el número más elevado de ASM institucionales, más difíciles de cumplir que los específicos por requerir el acuerdo entre áreas diferentes de la dependencia. En compensación, la Semarnat y la SEP tienen también el mayor número de ASM cumplidos, junto con CDI, en los dos últimos ciclos (Cuadro 38) y destacan por sus numerosas cancelaciones y bajas en 2013-2014 (cuadros 39 y 40). Esto deja a la Sedesol en la peor situación por su acumulación de ASM incumplidos.

En cuanto a la utilización instrumental en el GDF sólo hay datos disponibles del Evalúa DF en 2009-2011. Éstos muestran que el número de recomendaciones emitidas siempre fue superior al de las propuestas por los evaluadores y que se alcanzó una aceptación de las recomendaciones del 90% y un porcentaje de cumplimiento de sólo 72% de las aceptadas (65% de las emitidas) en el último año, a pesar del carácter vinculatorio del Evalúa DF.

Comparando situaciones, mientras el Coneval reporta haber alcanzado un igual nivel de cumplimiento de sus ASM del 82%, Evalúa DF informa 72% de sus recomendaciones aceptadas, pero existen varias diferencias de fondo: *a*) el Coneval acuerda los ASM con los operadores (éstos toman parte en la decisión de lo que pueden y están dispuestos a hacer), mientras Evalúa DF las decide, envía y espera su cumplimiento con base en su carácter vinculatorio, *b*) los ASM del primer caso sólo provocan reorientaciones de programas en 22% de los casos y las recomendaciones de rediseño del Evalúa DF constituyen 50%, y *c*) el nivel de transformación de las recomendaciones en ASM es muy bajo en el Coneval (45% en promedio en 2012-2013), mientras que es superior al cien por ciento en Evalúa DF.

Considerando este conjunto de mediciones, se concluye que ambos Consejos, a pesar de enfrentar fuertes resistencias burocráticas, han alcanzado un nivel relativamente importante de cumplimiento cuantitativo por caminos diferentes, quedando pendiente la evaluación del impacto que efectivamente han tenido en

las políticas y programas y fundamentalmente, en sus beneficiarios. Además, se debe cuidar de mantener los esfuerzos para aumentar el nivel de logro, cuantitativo y cualitativo, mientras en paralelo se trabaja en generar a largo plazo una cultura internalizada de evaluación.

Respecto de la utilización presupuestal, se ha mostrado que a nivel federal y del GDF no se percibe incidencia de las evaluaciones en la asignación de recursos con base tanto en los casos mexicanos analizados como en el estudio efectuado por Székely (2013) que recoge la opinión relativa al tema de una muestra de funcionarios.

Finalmente, en torno a la utilización conceptual se ha manifestado falta de estudios y dificultades metodológicas para identificar y medir avances; sin embargo, existen ejemplos de avances, tanto en el proceso del Coneval (i.e. inclusión de la evaluación de diseño a través de la MIR), como en el del Evalúa DF (i.e. diferenciación de enfoques y prácticas de evaluación a partir de la experiencia del Coneval), que podrían considerarse utilidades conceptuales. Para este tipo de utilización, un prerequisite fundamental lo constituye el acceso pleno a la información generada, especialmente por las evaluaciones, y los argumentos que conducen a las decisiones de los Consejos, a lo cual ha contribuido considerablemente la legislación en lo que va del siglo, pero aún se perciben algunas resistencias institucionales, sobre todo en materia de información financiera.

Describir la situación como ha sido presentada constituye un avance hacia su diagnóstico, sin embargo no basta, es necesario identificar también sus causas para tratar de incidir sobre éstas y es aquí donde la revisión del primer capítulo debe mostrar su utilidad. Durante mucho tiempo se ha trabajado en mejorar la calidad de las evaluaciones, suponiendo que eso provocaría su mayor utilización, pero resulta que, tanto la teoría como la práctica muestran que ese vínculo no es mecánico, ya que recomendaciones bien fundamentadas se pueden quedar “guardadas” si no convienen a sus operadores o a los diputados. Por otra parte, tampoco es de nuestro interés incrementar cualquier tipo de recomendaciones, sino las correctamente obtenidas para que redunden en mayor beneficio social. En consecuencia, se necesita garantizar que sólo las “buenas” recomendaciones sean aplicadas, lo que implica continuar trabajando, primero en lograr su calidad aceptable y luego en que sean utilizadas.

El tema de la calidad de las evaluaciones y sus múltiples determinantes organizativos y metodológicos (normatividad, recursos, planeación, formación y experiencia de los equipos, técnicas de investigación, actitud de los funcionarios, etcétera) ya ha sido ampliamente tratado.¹ En un contexto de recursos abundantes, amplia cultura y formación en evaluación tal vez pueda confiarse en que su calidad se logre sin necesidad de controles, pero en nuestros incipientes sistemas todavía es muy importante garantizarla. Para ello se necesita, además de contar con una combinación del mejor nivel posible de los factores determinantes señalados y acompañar el proceso desde los Consejos para ir resolviendo los problemas que se presenten y buscar sinergias, valorar la calidad obtenida en el informe final antes de proponer que sus recomendaciones se apliquen. Esa valoración se logra a partir de la metaevaluación, realizada por un experto contratado para ello o por los mismos Consejos. Para hacerla posible se necesitan diversos recursos: formación en metaevaluación para tener claridad sobre sus objetivos y métodos y tiempo para realizarla por los propios consejeros o dinero para contratarla externamente.

Como se mencionó en la segunda parte, en 2008-2012 Evalúa DF realizó distintos tipos de metaevaluaciones para garantizar la calidad, que implicaron conocimiento y sobre todo tiempo de trabajo; sin embargo, fueron posibles porque el número máximo de políticas evaluadas fue de tres y el de programas, de 13. Pero este mecanismo resulta muy difícil de reproducir cuando el número de evaluaciones que componen un ciclo se cuenta por cientos como sucede en Coneval, de ahí el riesgo de que los aumentos en cantidad se acompañen de reducciones en calidad. Ante esta situación sólo caben dos posibilidades: que los diputados modifiquen la LGDS para que no exija evaluaciones anuales de todos los programas o que Coneval organice metaevaluaciones aplicadas a una muestra de evaluaciones, con el alto riesgo de que las no evaluadas no cumplan con el requisito de calidad esperado.

Suponiendo que este primer problema logre resolverse, corresponde analizar el conjunto de factores que pueden incidir en la utilización, especialmente instrumental y presupuestal, de las recomendaciones cuya calidad ya haya sido reconocida. No hay duda de que todos los mencionados en el primer capítulo

¹ Al respecto véase Cardozo (2006 y 2012a).

de esta parte pueden aportar elementos a la explicación *ex post* de la utilización lograda. Sin embargo, cuando se busca incidir *ex ante* en la misma, es necesario diferenciar dos grupos de factores de la extensa lista resumida en el Cuadro 33: *a*) los que consideramos contextuales o exógenos porque el sistema actual de evaluación no los puede modificar, especialmente en el corto plazo, y *b*) los que son manipulables por dicho sistema en el corto plazo, permitiendo incrementar la probabilidad de utilización de una evaluación concreta.

De esta forma, el contexto político y legal, las relaciones intergubernamentales o el nivel de institucionalización de la cultura de evaluación afectan su grado de utilización (i.e. caso de Canadá), pero nuestros sistemas de evaluación no pueden modificarlos sustantivamente en el corto plazo, lo que tampoco significa que no se trabaje en ellos de forma paralela. En cambio, al interior del sistema de evaluación pueden utilizarse incentivos monetarios y simbólicos, individuales y colectivos, compromisos negociados y aplicación más efectiva de sanciones (i.e. Chile y Costa Rica) que impulsen la utilización de resultados de evaluación. La combinación de incentivos, al estilo chileno, es siempre más motivante que los mecanismos para evitar sanciones. En el caso del Evalúa DF, una de las fallas de su funcionamiento se debió a que el sistema de sanciones previsto no funcionó y que apenas se llegó a utilizar un incentivo simbólico de poco peso (oficio de reconocimiento por el cumplimiento de todas las recomendaciones). En el caso federal no hay sanciones puesto que los ASM no son vinculatorios, pero tampoco existe evidencia de la efectividad de los premios federales para promover mejores prácticas.

Pero más allá de los incentivos que han sido poco utilizados en el contexto mexicano, hay otros factores manipulables como: *a*) favorecer la participación de los responsables de los programas desde la decisión del tipo y contenido de la evaluación y la planeación conjunta de su posterior utilización hasta la discusión de las conclusiones, *b*) la realización de concursos abiertos y la selección de evaluadores prestigiados por su conocimiento y probidad, *c*) el levantamiento de percepciones de usuarios y organizaciones sociales vinculadas al tema; *d*) la presentación oportuna, relevante y accesible de la información; y *e*) la amplia difusión de informes y el monitoreo de su uso.

En relación con la participación de los responsables de los programas, cuyo involucramiento es altamente reconocido en la literatura, en México se ha interpuesto un obstáculo. Cuando la evaluación es solicitada por sus mismos

usuarios, como sucede habitualmente en otros contextos (Estados Unidos o Reino Unido), su integración en el proceso es casi natural; pero, cuando ésta es demandada por un organismo externo como el Coneval o Evalúa DF, pueden presentarse dos posibilidades: que el sistema confíe en el interés y colaboración del evaluado en un proceso de aprendizaje conjunto, o que considere que existe una alta probabilidad de que se resista, al menos en forma implícita, y entorpezca el proceso por ser ajeno a sus intereses. En el segundo caso es lógico pensar que no habrá apertura a la participación y esto es lo que parece haber estado sucediendo en México y otros países de América Latina. Nuevamente, la modificación de esta situación pasa a mediano plazo por cursos de actualización, espacios de discusión y publicaciones accesibles a los funcionarios y, en lo inmediato, por la adopción de un sistema de incentivos.

La realización de concursos, la selección de evaluadores y la participación de usuarios y organizaciones sociales empiezan por afectar la calidad de los trabajos (vía su rigurosidad metodológica y su amplitud de puntos de vista) pero también su utilización a partir de la credibilidad, tantas veces señalada por los autores revisados, especialmente a nivel conceptual y por la sociedad misma. Si todas las instituciones de prestigio pueden participar en el proceso y los criterios de selección son explícitos y aplicados en igualdad de condiciones, sus resultados resultan confiables pero, si además se ha tenido la oportunidad de participar y conocer de primera mano un buen trabajo realizado, la credibilidad tiende a aumentar. Esto se refuerza también con una amplia difusión del informe. Se trata de temas que no se discuten en el discurso mexicano, pero que en la práctica se aplican cada vez menos, entre otras razones porque implican mayores costos, especialmente de llevar a cabo los trabajos de campo, ya casi inexistentes.

Además de la difusión del informe, la presentación de la información en un lenguaje atractivo y accesible al lector así como su extensión limitada son determinantes para facilitar su lectura y posterior utilización. Por esta razón se ha sugerido que, además del resumen ejecutivo que en la mayor parte de los casos mexicanos se dispone, se intenten otras versiones dirigidas a públicos diferentes (funcionarios, legisladores, investigadores, ciudadanía en general). Otro aspecto importante es su oportunidad para que pueda ser tomada en cuenta en procesos con plazos estrictos como puede ser la aprobación del PEF, en cuya adecuación se han hecho avances sin que el problema se haya resuelto plenamente.

En cuanto al monitoreo de la utilización del informe y sus recomendaciones, éste constituye un mecanismo de coerción sobre las dependencias y entidades responsables de realizarlo, propio del supuesto de que los funcionarios no lo hacen voluntariamente. En consecuencia, si el control se diluye, la posibilidad de utilización también se reduce. Además, el monitoreo permite a los investigadores y público interesado conocer los avances del sistema. Su aplicación es responsabilidad de los Consejos de evaluación y dista mucho de estar actualizado y ser útil: Evalúa DF no lo ha publicado desde mediados de 2011 y Coneval, ante la cantidad de información manejada, sólo publica un informe anual agregado por dependencias y entidades que no permite el seguimiento por programas, su redacción no siempre es clara, y dificulta su seguimiento año con año.

Por último, la utilización presupuestal ha sido importante en países en que el propio Poder Legislativo solicita o realiza evaluaciones (Estados Unidos, Reino Unido) y muestra avances en el caso chileno. Esta es la coerción probablemente más fuerte que puede pesar en la utilización instrumental por la administración pública. Lograrla depende también del contexto político y la cultura y educación de los diputados al respecto, así como de la ya mencionada oportunidad del informe, pero incentivarla se muestra más difícil que en el caso de los funcionarios. En consecuencia, la posibilidad de que la información generada por la evaluación sea tomada en cuenta en nuestro contexto es más probable a nivel de la SHCP, encargada de realizar el proyecto de presupuesto; sin embargo, sería conveniente sensibilizar más a sus funcionarios con la información social, no sólo con la económico-financiera que están habituados a manejar.

Síntesis general de conclusiones

Myriam Cardozo Brum

Desarrollar e internalizar una cultura de la evaluación en diferentes grupos sociales (funcionarios, diputados, investigadores, organizaciones no gubernamentales) y la ciudadanía en general constituye la máxima aspiración de quienes estamos convencidos de su necesidad y conveniencia, para que ésta sea voluntariamente aplicada, desarrollada en forma colaborativa, sus productos sean de alta calidad y sus resultados ampliamente utilizados. Pero lograrlo requiere tiempos muy prolongados de institucionalización, como ha sucedido en Estados Unidos de América a pesar de ser el único país revisado que cuenta tanto con el compromiso de su Poder Ejecutivo como del Legislativo. Los procesos han avanzado en forma más lenta en Reino Unido y Canadá, mientras que los países latinoamericanos que han iniciado recientemente su institucionalización formal (Colombia, Chile, México) tienen aún un largo camino por recorrer, que Chile está encarando con la mayor velocidad. En los tres casos mencionados se ha otorgado un alto valor al enfoque experimental y econométrico, al igual que en Estados Unidos de América, aunque no siempre ha sido posible aplicarlo por restricciones técnicas y financieras. También se ha cuidado poco la realización de metaevaluaciones que contribuyan a mejorar la calidad de los trabajos, preocupación especialmente cuidada por los canadienses.

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 estos tres aspectos vuelve a sobresalir Estados Unidos de América, mientras Francia lo hace por su profusa publicación de libros y los múltiples debates generados en sus sociedades de evaluación regionales, además de la nacional, Canadá por sus diversos programas de formación universitaria y preocupación por la calidad de los trabajos, y Reino Unido por haber aglutinado la reflexión al funcionar como sede de la sociedad internacional que agrupa fundamentalmente a las naciones europeas. En torno a estos temas, los países de América Latina se caracterizan por sus grandes limitaciones.

México realizó su primer esfuerzo de amplitud en los años 2001-2006; su trabajo se concentró en la Sedesol y, en menor grado, en la Sagarpa; mientras que las otras secretarías de Estado (Educación y Salud principalmente, Medio Ambiente en forma más limitada) se han ido sumando al proceso después de la creación del Coneval en 2006. Esto ha provocado un notable aumento del número de evaluaciones realizadas y la simplificación de los trabajos, lo que lleva a cuestionarse sobre la calidad de los mismos, tema tratado en la segunda parte del libro mediante la aplicación de múltiples metaevaluaciones.

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 muchos de ellos carecen de solidez metodológica y de aportaciones (EED, de igual manera entre las frecuentes), lo que no siempre se explica por la limitación de sus recursos. En sus procesos suelen detectarse problemas como: falta de TR o disponibilidad sólo de términos genéricos; adjudicaciones directas de contratos no únicamente a centros de educación superior, de investigación científica u organizaciones no lucrativas, como marca el artículo 73 de la LGDS, sino también a consultoras; 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 anterior contrasta con la vivida a nivel federal en 2001-2006 en que, aunque existían muchos problemas (normatividad inadecuada, falta de experiencia, ausencia de evaluaciones de diseño, etcétera), se hacía un gran esfuerzo por llevar a cabo evaluaciones integrales, con trabajo de campo e informes más amplios. También con la evaluación realizada en el GDF en 2008-

2014 que cuenta siempre 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 han 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 (i.e. INEE en educación).

En síntesis, la situación federal actual muestra una tendencia a cumplir con el requisito formal de realizar una evaluación anual por programa, cuya calidad no está garantizada¹ y, en consecuencia, tampoco la conveniencia de utilizar sus recomendaciones, que se revisa en la tercera parte de este libro.

Se analizaron los diversos tipos de utilización de resultados posibles en el contexto mexicano federal y del GDF: instrumental (la más desarrollada), presupuestal y conceptual. 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 en la capital del país ya lo ha puesto en práctica con resultados no tan elevados como se esperaba debido a la poca efectividad de las sanciones previstas.

La teoría y las experiencias mundiales sobre utilización de resultados de evaluación se concentran también en la de tipo instrumental, que realizan los propios evaluados. Muestran un sinnúmero de factores que pueden explicar el limitado nivel 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, en el nuestro, no nos interesa promover la utilización al margen de ella. 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,

¹ En este trabajo no se revisó el tema de los determinantes de la calidad debido a que ya han sido cubiertos en trabajos previos (Cardozo, 2006, 2012a y b).

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.

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 como su posterior utilización, para que los esfuerzos realizados cobren el sentido que justifica la existencia del trabajo evaluativo en nuestro contexto.

Referencias

Libros y revistas

- Aberlen, E. (2014), *L'évaluation en évolution. Pratiques et enjeux de l'évaluation dans le contexte de l'efficacité du développement*, Collection Conférences et Séminaires. France: Agence Française du développement.
- Álvarez, J. y L. Niño (2007), *Políticas educativas. evaluación y metaevaluación*. Bogotá, Colombia: Universidad Pedagógica Nacional.
- Anderson, L. et al. (2006), *Evaluación de programas: estudios de gran escala y de pequeña. Algunas generalizaciones*, Serie políticas educativas, núm. 5. México: INEE.
- Appleton-Dyer, S. et al. (2012), "Understanding evaluation influence within public sector partnerships", *American Journal of Evaluation*, vol. 33(4), pp. 532-546.
- Arellano, D. (2012), *Sistemas de evaluación del desempeño para organizaciones públicas*. México: CIDE.
- Association France Qualité Publique (2010), *Acteurs de la qualité et de l'efficience publiques. Guide Pratique*. France: La Documentation française.
- Ballart, X. (1993), "Evaluación de políticas. Marco conceptual y organización institucional", *Revista de Estudios Políticos* (nueva época), núm. 80, pp. 199-224.
- Balthasar, A. (2009), "Institutional design and utilization of evaluation. A contribution to a theory of evaluation influence based on Swiss experience", *Evaluation Review*, vol. 33(3), pp. 226-256.
- Barbier, J.C. (2012), "Évaluations expérimentales: quelques leçons de l'histoire", *Politiques sociales et familiales*, núm. 110. France: Centre d'économie de la Sorbonne / Université Paris 1 Panthéon Sorbonne.
- Baron, G. y N. Matyasik (2012), *L'évaluation des politiques publiques: défi d'une société en tension*. París: L'Harmattan.
- BID (2010), *La gestión para resultados en el desarrollo: avances y desafíos en América Latina y el Caribe*. Estados Unidos.

- Banco Mundial (2004), *La pobreza en México: una evaluación de las condiciones, las tendencias y la estrategia del gobierno*. México.
- Bannon, Y. *et al.* (2009), *Manuel du responsable en évaluation de programme*. Quebec, Canadá: GRAEP.
- Barbier, J.C. (2012), “Évaluations ‘expérimentales’: quelques leçons de l’histoire”, *Politiques Sociales et Familiales*, núm. 110, pp. 19-31.
- Barrington, G.V. (2012), *Consulting start-up and management: a guide for evaluators and applied researchers*. Thousand Oaks, CA: Sage.
- Baslé, M. (2008), *Économie, conseil et gestion publique : suivi et évaluation des politiques publiques et des programmes*. París: Economica.
- Battesti, J.P. *et al.* (2012), *Cadrage méthodologique de l’évaluation des politiques publiques partenariale: guide*. París: IGA/IGF/IGAS.
- Blechinger, P. *et al.* (2011), “A multi-criteria evaluation of policy instruments for climate change in the power generation sector of Trinidad and Tobago”, *Energy Policy*, núm. 39, pp. 6331-6343.
- Boaz, A. y C. Hayden (2002), “Pro-active evaluators: enabling research to be useful, usable and used”, *Evaluation*, vol. 8(4), pp. 440-453.
- Bonbright, D. (2012), *Use of impact evaluation results*. Estados Unidos: Interaction & The Rockefeller Foundation.
- Bouquet, A. (2013), *L’évaluation des politiques publiques: états de l’art et controverses*. Strasbourg /París: ENA.
- Bozio, A. y J. Grenet (coord.) (2010), *Economie des politiques publiques*, Collection Repères Economie, núm. 554, France: La Découverte.
- Braconier, P. y G. Cauquil (2010), *L’évaluation des politiques publiques: le développement d’une nouvelle culture*. París: SCEREN/CNDP.
- Bustelo, M. (2003), “¿Qué tiene de específico la metodología de la evaluación?”, en Bañón (comp.) *La evaluación de la acción y de las políticas públicas*. Madrid: Díaz de Santos.
- (2004), *La evaluación de políticas públicas de igualdad de género en España*. Madrid: Catarata.
- 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/Miguel Ángel Porrúa.
- (2009), *Evaluación y metaevaluación en los programas mexicanos de desarrollo social*, Colección Teoría y Análisis. México: UAM-Xochimilco.
- (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, pp. 43-59.

- (2012a), *Evaluación y metaevaluación en las políticas y programas públicos. El estado del arte*. México: UAM-Xochimilco.
- (2012b), *Sistematización de la experiencia de evaluación de Evalúa DF en 2008. Metaevaluación y autoanálisis*. México: UAM-Xochimilco/Evalúa DF.
- (2013), “De la evaluación a la reformulación de políticas públicas”, *Política y Cultura*, núm. 40, México: UAM-Xochimilco, pp. 123-149.
- Cardozo, M. y A. Mundo (2012), *Guía de orientación para la evaluación de políticas y programas de desarrollo social*. México: Evalúa DF.
- Castro, C. (2007), *Evaluación de políticas públicas. El Programa Supera en San Luis Potosí*. México: INAP/Copocyt.
- Cejudo, G. y C. Maldonado (eds.) (2011), *De las recomendaciones a las acciones*. México: CIDE/CLEAR/SFP.
- Cifuentes, M.A. et al. (2000), *Medición de la efectividad del manejo de áreas protegidas*. Costa Rica: BMZ/GTZ/WWF/UICN.
- Contandriopoulos, D. y A. Brousselle (2012), “Evaluation models and evaluation use”, *Evaluation*, vol. 18(1), pp. 61-77.
- Cooksy, L. (2005), “Quality, context and use issues in achieving the goals of metaevaluation”, *American Journal of Evaluation*, vol. 26, núm. 1, pp. 31-42.
- Cooksy, L.J. y V. Caracelli (2005), “Quality, context, and use: issues in achieving the goals of metaevaluation”, *American Journal of Evaluation*, vol. 26, núm. 1, pp. 31-42.
- (2009), “Metaevaluation in practice selection and application of criteria”, *Journal of Multidisciplinary Evaluation*, vol. 6, núm. 11, pp. 1-15.
- Cortés, F. et al. (2008), *Método científico y política social. A propósito de las evaluaciones cualitativas de programas sociales*. México: El Colegio de México.
- Cousins, J.B. y K. Leithgood (1986), “Current empirical research on evaluation utilization”, *Review of Educational Research*, núm. 56, pp. 331-364.
- Cousins, J.B. y J. A. Chouinard (2012), *Participatory evaluation up close: an integration of research-based knowledge*. Charlotte, NC: Information Age Publishing.
- Coryn, Ch. (2004), “The state of evaluation in Canada”, *Journal of MultiDisciplinary Evaluation*, vol. 1, núm. 1, pp. 55-68.
- (2005), “An update on evaluation in Canada”, *Journal of MultiDisciplinary Evaluation*, vol. 2, núm. 2, pp. 130-131.
- (2005), “Evaluation in Canada”, *Journal of MultiDisciplinary Evaluation*, vol. 2, núm. 3, pp. 166-171.
- Crabbé, A. y P. Leroy (2008), *The handbook of environmental policy evaluation*. Londres: Earthscan.

- Cunill, N. y S. Ospina (2003), *Evaluación de resultados para la gestión pública moderna y democrática: experiencias latinoamericanas*. Venezuela: CLAD/AECI/MAP/FIIAPP.
- (2008), *Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina. Informe comparativo de 12 países*. Banco Mundial/CLAD.
- Custodio, E. (2002), “Aquifer overexploitation: what does it mean?”, *Hydrogeology Journal*, vol. 10, núm. 2, pp. 254-277.
- Daigneault, P. (2010), “L'examen de la qualité des évaluations fédérales: une méta-évaluation réussie?”, *Canadian Journal of Program Evaluation*, vol. 23, núm. 2, pp. 191-224.
- Díaz, L.F. (2002), “La metaevaluación y su método”, *Revista de Ciencias Sociales*, año XLIII, núms. 92/93. Costa Rica: Universidad de Costa Rica.
- (2005), “Metaevaluación del modelo del Sistema Nacional de Evaluación de Costa Rica”, ponencia presentada en el *X Congreso Internacional del CLAD*. Santiago de Chile.
- Díaz, L.F. y R. Rosales (2003), *Metaevaluación: evaluación de la evaluación de políticas, programas y proyectos sociales*. San José de Costa Rica: Editorial Universidad Estatal a Distancia.
- Donaldson, S.I. *et al.* (2009), *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage.
- Dufló, E. y A. Banerjee (2009), “L'approche expérimentale en économie du développement”, *Revue d'Économie Politique*, vol. 115, núm. 9, pp. 692-726.
- Egaña, R. (2008), *Consideraciones sobre un diagnóstico de sistemas de M&E*. Santiago de Chile: CLAD.
- Feinstein, O. (2002), “Use of evaluations and the evaluation of their use”, *Evaluation*, vol. 8(4), pp. 433-439.
- Feinstein, O. y G. Hernández Licona (2008), *El papel de la evaluación en México: logros, desafíos y oportunidades*. México: SHCP/BM.
- Fetterman, D.M., S. Kaftarian y A. Wandersman (2014), *Empowerment evaluation: knowledge and tools for self-assessment, evaluation capacity building, and accountability*. Thousand Oaks, CA: Sage.
- Figueroa, F. y V. Sanchez-Cordero (2008), “Effectiveness of natural protected areas to prevent land use and land cover change in Mexico”, *Biodiversity Conservation*, núm. 17, pp. 3223-3240.
- Forss, K. *et al.* (2002), “Process use of evaluations”, *Evaluation*, vol. 8(1), pp. 29-45.
- Forss, K., M. Marra y R. Schwartz (eds.) (2011), *Evaluating the complex. Attribution, contribution and beyond, Comparative Policy Evaluation*, vol. 18. New Brunswick (USA)/ Londres (UK): Transaction Publishers.
- Frenk, J. (2012), “La salud como derecho ciudadano”, *Nexos*, pp. 40-50.

- Funnell, S. y P. Roger, P. (2011), *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco: Jossey-Bass.
- Fouquet, A. y L. Méasson (2009), *L'évaluation des politiques publiques en Europe: culture et futurs*, Société Française de l'Évaluation-*Revue Française de Socio-Economie*. France: L'Harmattan.
- García, E. (2009), "Metaevaluación de políticas públicas: una visión desde la ciencia política", *Revista Reforma y Democracia*, núm. 43. Venezuela: CLAD.
- García, R. y M. García (2010), *La gestión para resultados en el desarrollo. Avances y desafíos para América Latina y el Caribe*. BID.
- Gargani, J. (2011), "More than 25 years of the American Journal of Evaluation: the recollections of past editors in their own words", *American Journal of Evaluation*, núm. 32(3), pp. 428-447.
- Garrido, F. y S. López (2011), "Evaluación de los programas y servicios de salud en México", *Revista de Salud Pública de México*, vol. 53, Supl. 4, s399-s401.
- Gómez, O. *et al.* (2004), "La equidad y la imparcialidad en la reforma del Sistema Mexicano de Salud", *Revista de Salud Pública de México*, vol. 46(5), pp. 399-416.
- (2011), "Sistema de salud de México", *Revista de Salud Pública de México*, vol. 53(2), pp. 220-232.
- Greener, I. y B. Greve (2014), *Evidence and evaluation in social policy*. Chichester, West Sussex: Wiley Blackwell.
- Guba, E.G. (1975), "Problems in utilizing the results of evaluation", *Journal of Research and Development in Education*, vol. 8(3), pp. 42-54.
- Guba, E. y Y. Lincoln (1989), *Fourth generation evaluation*. Newbury Park: Sage Publications.
- Hanssen, C. *et al.* (2008), "Concurrent meta-evaluation: a critique", *American Journal of Evaluation*, vol. 29, núm. 4, pp. 572-582.
- Hedler, H. y N. Gibram (2009), "The contribution of metaevaluation to program evaluation: proposition of a model", *Journal of Multidisciplinary Evaluation*, vol. 6, núm. 12, pp. 210-223.
- Hockings, M. *et al.* (2000), *Evaluating Effectiveness. A framework for Assessing the management of Protected Areas*. Suiza y Reino Unido: IUCN.
- Højlund, S. (2014), "Evaluation use in the organizational context –changing focus to improve theory", *Evaluation*, vol. 20(1), pp. 26-43.
- Holden, D. y M. Zimmerman (2009), *A practical guide to program evaluation planning*. Thousand Oaks, CA: Sage.
- House, E.R. (1997), *Evaluación, ética y poder*. Madrid: Morata.

- Howlet, M. (2009), "Policy analytical capacity and evidence-based policy-making: Lessons from Canada", *Canadian Public Administration*, vol. 52, Issue 2: pp. 153-175.
- Huitema, D. *et al.* (2008), *Climate change policy evaluation across Europe: Practices of climate policy evaluation in Finland, France, Germany, Poland, Portugal, the UK and the European Union*. Holanda: Institute for Environmental Studies.
- Hum, D. y W. Simpson (2002), "Public training programs in Canada: A meta-evaluation", *Canadian Journal of Program Evaluation*, vol. 17, núm. 1, pp. 119-138.
- Jacob, S. (2006), "Trente ans d'évaluation de programme au Canada: l'institutionnalisation interne en quête de qualité", *Revue Française d'Administration Publique*, núm. 119, pp. 515-531.
- Jacob, S. y G. Desautels (2014), "Assessing the quality of aboriginal program evaluations", *Canadian Journal of Program Evaluation*, vol. 29, núm. 1, pp. 62-86.
- Jacobson, C. *et al.* (2011), "Maximizing conservation evaluation utilization", *Evaluation*, vol. 17(1), pp. 53-71.
- Jiménez, P. (2003), "La evaluación de políticas públicas en los Estados Unidos y el papel de la General Accounting Office", *Auditoría Pública*, núm. 31, pp. 34-43.
- Johnson, K. *et al.* (2009), "Research on evaluation use: A review of the empirical literature from 1986 to 2005", *American Journal of Evaluation*, vol. 30(3), pp. 377-410.
- Joint Committee on Standards for Educational Evaluation (1994), *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.
- King, J.A. y L. Stevahn (2013), *Interactive evaluation practice: mastering the interpersonal dynamics of program evaluation*. Thousand Oaks, CA: Sage.
- Kleiman D.G. *et al.* (2000), "Improving the evaluation of conservation programs", *Conservation Biology*, núm. 14(2), pp. 356-365.
- Kumar, K. (2013), *Evaluating democracy assistance*. Boulder, CO: Lynne Rienner Publishers.
- LaBelle, M. (2009), "Metaevaluation as a means of examining evaluation influence", *Journal of Multidisciplinary Evaluation*, vol. 6, núm. 11, pp. 32-37.
- Lacasse, M.F. (1993), "Gestion et émasculature des évaluations: le cas de Canada", *Revue Française d'Administration Publique*, núm. 66, pp. 274-276.
- Lascoumes, P. (1998), "Pratiques et modèles de l'évaluation" y "Usages et retombées de l'évaluation", en M.Ch. Kesler *et al.* *Evaluation des politiques publiques*. París: L'Harmattan.
- Laubli, M. y J. Mayne (eds.) (2014), *Enhancing evaluation use: insights from internal evaluation units*. Estados Unidos: Sage Publications.
- LaVelle, J. y S. Donaldson (2010), "University-based evaluation training programs in the United States 1980-2008: An empirical examination", *American Journal of Evaluation*, núm. 31(1), pp. 9-23.

- Le Ster, C. (2011), *Revue de littérature sur les méthodologies de l'évaluation* [www.entrepreneursdumonde.org], fecha de consulta : 8 de septiembre de 2014.
- Leviton, L. (2003), "Evaluation use: advances, challenges and applications", *American Journal of Evaluation*, vol. 24(4), pp. 525-535.
- Leviton, L. y E. Hughes (1981), "Reserch on the utilization of evaluations", *Evaluation Review*, vol. 5, núm. 4, pp. 525-548.
- Magaña, R. y G. Vargas (coords.) (2004), *Evaluación de las políticas públicas: redefinición o continuidad*. México: UAM-Iztapalapa.
- Maldonado, C. y C. Galíndez (2013), *Monitoreo, evaluación y gestión por resultados*. México: CIDE/CLEAR.
- Martínez Rizo, F. y E. Blanco (2010), "La evaluación educativa: experiencias, avances y desafíos", en A. Arnaut y S. Giorguli (coords.), *Los grandes problemas de México. Educación*. México: Colegio de México, pp. 89-123.
- Mayne, J. (1994), "Utilizing evaluation in organizations: the balancing act", en F. Leeuw, R. et al. (eds.), *Can governments learn? Comparative perspectives on evaluation and organizational learning*. New Brunswick, New Jersey: Transaction Publishers.
- Mc David, J. et al. (2013), *Program evaluations and performance measurement*. Canadá: University of Victoria.
- Mejía, J. (2003), *La evaluación de la gestión y las políticas públicas*. México: Miguel Ángel Porrúa.
- Mokate, K. (2003), *Convirtiendo al "monstruo" en aliado: la evaluación como instrumento de gerencia social*. Washington, DC: INDES/BID.
- Monnier, E. (1995), *Evaluación de la acción de los poderes públicos*. España: Instituto de Estudios Fiscales.
- Montiel, M.J. (2011), "La evaluación como instrumento de cambio. ¿Cómo se utilizan las evaluaciones en los programas gubernamentales en México?", en G. Cejudo y C. Maldonado (eds.), *De las recomendaciones a las acciones*. México: CIDE/CLEAR/SFP.
- Morell, J. (2010), *Evaluation in the face of uncertainty: anticipating surprise and responding to the inevitable*. Nueva York: Guilford Publications.
- Mouterde, F. y S. Trosa (2010), *Les nouvelles frontières de l'évaluation 1989-2009: vingt ans d'évaluation des politiques publiques en France, et demain?* France: L'Harmattan.
- Murnane, R.J. et al. (2006). "¿Ha contribuido el Programa Escuelas de Calidad a mejorar la educación pública en México?", en Reimers, F. (ed.), *Aprender más y mejor; políticas, programas y oportunidades de aprendizaje en educación básica en México 2000-2006*. México: Fondo de Cultura Económica.

- Muthiga, N.A. (2009), "Evaluating the effectiveness of management of the Malindi-Watamu marine protected area complex in Kenya", *Ocean & Coastal Management*, núm. 52, pp. 417-423.
- Nkwake, A., Trandafil, H. y Hughey, J. (2013), "Examining the relationship between community participation and program outcomes in a metaevaluation", *Journal of Multidisciplinary Evaluation*, vol. 9. núm. 20, pp. 1-17.
- Offredi, C. y F. Ravoux (2010), *La notion d'utilité sociale au défi de son identité dans l'évaluation des politiques publiques*. París: l'Harmattan.
- Olavarría, M. (2012), "La evaluación de programas en Chile: análisis de una muestra de programas públicos", *Reforma y Democracia*, núm. 54, pp. 1-21.
- OCDE (2011). *Análisis del extensionismo agrícola en México. 50 mejores políticas para una vida mejor*. París.
- Parry, M. *et al.* (1998), "Adapting to the inevitable", *Nature*, vol. 395, núm. 6704, pp. 741-742.
- Patton, M.Q. (1978), *Utilization focused evaluation*. Thousand Oaks: Sage Publications.
- (1998), "Discovering process use", *Evaluation*, vol. 4 (2), pp. 225-233.
- (2011), *Developmental evaluation: applying complexity concepts to enhance innovation and use*. Nueva York: Guilford Publications
- (2011), *Essentials of utilization focused evaluation*. Thousand Oaks: Sage Publications.
- (2012), "A utilization-focused approach to contribution analysis", *Evaluation*, vol. 18(3), pp. 364-377.
- (2013), "Meta-evaluation: Evaluating the evaluation of the Paris Declaration", *Canadian Journal of Program Evaluation*, vol. 27, núm. 3, pp. 147-171.
- Perret, B. (2008), *L'évaluation des politiques publiques*. París: La Découverte.
- Pielke, J.R. *et al.* (2007), "Climate change 2007: Lifting the taboo on adaptation", *Nature*, vol. 395, núm. 7128, pp. 597-598.
- Pomeroy R. *et al.* (2004), *How is your MPA doing? A guidebook of natural and social indicators for evaluating marine protected area management effectiveness*. Suiza y Reino Unido: IUCN/WWF/NOAA.
- Pontones, C. y R. Pérez (2013), "El value for money en los servicios sociales: el caso de Nottinghamshire", *CIRIEC-España, Revista de Economía Pública, Social y Cooperativa*, núm. 79, pp. 145-165.
- Ramos, J.M. (2012), *La creación de capacidades institucionales en materia de evaluación: la experiencia mexicana (2007-2012)*. México: El Colef-Iapem.
- (2013), *Evaluación del desempeño: gestión y política para la generación de resultados efectivos*. México: Auditoría Superior de la Federación.

- Ramos, J.M. *et al.* (coords, 2011), *La evaluación de las políticas públicas en México*. México: INAP/Colegio de la Frontera Norte.
- Revista Ciencias Políticas y Administración Pública* [http://www.ucecentral.cl/prontus_ucecentral2012/site/edic/base/port/inicio.html], fecha de consulta: 25 de noviembre de 2014.
- Revista Iberoamericana de Evaluación Educativa* [<http://www.oei.es/noticias/spip.php?article2775>], fecha de consulta: 25 de noviembre de 2014.
- Revista INVI* [<http://revistainvi.uchile.cl/index.php/INVI/search/results>], fecha de consulta: 25 de noviembre de 2014.
- Revista Universum* [<http://www.utralca.cl/link.cgi/#/link.cgi/>], fecha de consulta: 25 de noviembre de 2014.
- Revue Cités* (2009), *Lidéologie de l'évaluation. La grande imposture*, Presses Universitaires de France núm. 37.
- Revue Française de Socio-Economie* (2010), *Les politiques de quantification*, Francia: La Découverte.
- Revue Mesure et Évaluation en Éducation*, Association pour le Développement de la Mesure et de l'Évaluation en Éducation [<http://www.admee.org/index.php/la-revue-mee>].
- Ridde, V. *et al.* (2003), "L'évaluation participative de type empowerment: une stratégie pour le travail de rue", *Revue Service Social*, vol. 50, núm. 1, pp. 263-279.
- Ridde V. y C. Dagenais (2009), *Approches et pratiques en évaluation de programmes*. Canadá: PU de Montréal.
- Rivero, C. y M. Gabaldón (1992), "Evaluación de sistemas de áreas naturales protegidas: una metodología numérica", *PARQUES*, vol. 3, núm. 1, Gland, Suiza: UICN.
- Rodríguez, J. (2006), "La Contraloría General de los Estados Unidos. Funcionamiento y auditorías", *Revista Española de Control Externo*, núm. 22, pp. 137-151.
- Rodríguez, L. e I. Espejel (en prensa), "Los criterios de sustentabilidad en los programas oficiales aplicados a las áreas naturales protegidas federales de México", *Investigación Ambiental. Ciencia y Política Pública*, INECC.
- Rossi, P.H. y H.E. Freeman (1993), *Evaluation: A systematic approach*. Newbury Oark: Sage Publications.
- Roth, A. (2009), "La evaluación de políticas públicas en Colombia: una mirada crítica a partir de las prácticas evaluativas oficiales de los programas de la Red de Apoyo Social", *Revista del CLAD Reforma y Democracia*, núm. 45, pp. 161-186.
- Ridde, V. y Ch. Dagenais (2009), *Approches et pratiques en évaluation de programmes*. Canadá: Les Presses de l'Université de Montréal.
- Rivera, M.G. y P. del Monte-Luna, (2011), "Estado de la investigación evaluativa en el caso de las áreas marinas protegidas de México", *Revista Oceanídes*, vol. 26, núm. 2, pp. 9-17.

- Russ-Eft, D. *et al.* (2008), *Evaluator competencies standards for the practice of evaluation in organizations*. San Francisco: Jossey-Bass.
- Ryan, K. y Cousins, J.B. (2009), *The SAGE International handbook of educational evaluation*. Thousand Oaks, CA: Sage Publications, Inc.
- Saint Clair, W. (1993), "L'évaluation des politiques publiques au Royaume Uni", *Revue Française d'Administration Publique*, núm. 66, pp. 265-269.
- Sanders, J.R. y D.N. Nafziger (2011), "A basis for determining the adequacy of evaluation designs", *Journal of Muldisciplinary Evaluation*, vol. 7, núm. 15, pp. 44-78 (originalmente publicado como Paper núm. 7, Occasional Paper Series, abril, 1976).
- Saunders, M. (2012), "The use and usability of evaluation outputs: A social practice approach", *Evaluation*, vol. 18(4), pp. 421-436.
- Schroeter, D.C. (2009), "Metaevaluation", en L.E. Sullivan (ed.), *The Sage Glossary of the Social and Behavioral Sciences*. Nueva York: Sage.
- Schwandt, T. y E. Halpern (1988), *Linking auditing and metaevaluation*. Newbury Park: Sage Publications.
- Schwentesius, R. *et al.* (2007), "Metaevaluación de tres evaluaciones oficiales de la Sagarpa del programa de pagos directos (Procampo) a la agricultura mexicana", *Estudios Sociales*, vol. 16, núm. 30, pp. 104-134.
- Scriven, M. (1967), "The methodology of evaluation", en Tyler *et al.*, *Perspective of curriculum evaluation*. Chicago: Rand McNally.
- (1969), "An introduction to meta-evaluation", *Educational Products Report*, núm. 2, pp. 36-38.
- (1973), "Goal free evaluation", en House, E. (ed.), *The politics and process*. Nueva York: McCutchan.
- (1976), "Evaluation bias and its control", *Evaluation Studies Review Annual*, vol. 1.
- (1991), *Evaluation Thesaurus*. Newbury Park: Sage Publications.
- (2009), "Meta-evaluation revisited", *Journal of Multidisciplinary Evaluation*, vol. 6, núm. 11, pp. iii-viii.
- Smith, M.J. (2010), *Handbook of program evaluation for social work and health professionals*. Oxford University Press.
- Spaulding, D.T. (2008), *Program evaluation in practice: core concepts and examples for discussion and analysis*. San Francisco: Jossey-Bass.
- Stake, R. (2011), "Program evaluation particularly responsive evaluation", *Journal of Muldisciplinary Evaluation*, vol. 7, núm. 15, pp. 180-201 (originalmente publicado como Paper núm. 5, Occasional Paper Series, noviembre, 1975).
- Steam, C. *et al.* (2005), "Monitoring and evaluation in conservation: A review of trends and approaches", *Conservation Biology*, núm. 19(2), pp. 295-309.

- Stufflebeam, D.L. (1974). Meta-evaluation. Paper núm. 3. Occasional Paper Series. Kalamazoo: Western Michigan University Evaluation Center.
- (2001), “The meta-evaluation imperative”, *American Journal of Evaluation*, vol. 22, núm. 2, pp. 183-209.
- (2009), “Metaevaluation”, *Journal of Multidisciplinary Evaluation*, vol. 7, núm. 15, pp. 99-158 (originalmente publicado como Paper núm. 3, Occasional Paper Series, diciembre, 1974).
- Stufflebeam, D. y A. Shinkfield (1993), *Evaluación sistemática*. España: Paidós.
- Stockmann, R. (2009), *Evaluación y desarrollo de la calidad. Base para una administración de la calidad enfocada hacia el impacto*. Costa Rica: Universidad de Costa Rica.
- (2011), *Manual de Evaluación. Una guía práctica de procedimientos*. Costa Rica: Universidad de Costa Rica.
- Stoll-Kleemann, S. (2010), “Evaluation of management effectiveness in protected areas: methodologies and results”, *Basic and Applied Ecology*, núm. 11, pp. 377-382.
- Tourgeon, J. (1998), “Le programme de l’action communautaire pour les enfants”, en M.Ch. Kessler et al., *Evaluation des politiques publiques*. París: L’Harmattan.
- Thoenig, J. C. (2005), “L’évaluation. Un cycle de vie à la française”, en F. Lacasse y P.E. Verrier (dir.), *30 ans de réforme de l’État. Expériences françaises et étrangères. Stratégies et bilans*. París: Dunod, pp. 117-127.
- Trosa, S. (2003), *L’évaluation des politiques publiques*. Francia: Institut de l’Entreprise.
- Valovirta, V. (2002), “Evaluation utilization as argumentation”, *Evaluation*, vol. 8(1), pp. 60-80.
- Vargas, G. (2007), *Metaevaluación de los programas de la Secretaría de Agricultura (Sagarpa), dirigida a productos agrícola básicos generales*. México: CEDRSSA.
- Vollet, D. y F. Hadjab (2008), *Manuel de l’évaluation des politiques publiques*. Versailles: Editions Quae.
- Weiss, C. (1972), *Utilization of evaluation: toward comparative study*. Boston, MA: Allyn and Bacon.
- (1988), “Evaluation for decisions: Is anybody there? Does anybody care?”, *Evaluation Practice*, vol. 9(1), pp. 5-19.
- Wiesner, E. (2000), *Función de evaluación de planes, programas, estrategias y proyectos*. Santiago de Chile: CEPAL.
- Wingate, L.A. (2010), “Metaevaluation: purpose, prescription, and practice”, en Baker, E. et al. (eds.), *International Encyclopedia of Education*, Elsevier.
- Worthen, B.R. y J.R. Sanders (2011), “Content specialization and educational evaluation: A necessary marriage”, *Journal of Multidisciplinary Evaluation*, vol. 7, núm. 15, pp. 272-295 (publicado como Paper núm. 14, Occasional Paper Series, septiembre, 1984).

- Yoshizawa, G. (2009), “Eight models of evaluations use”, *Journal Japanese of Evaluation Studies*, vol. 9(2), pp. 51-61.
- Zaltman, A. (2008), *Experiencias con la institucionalización de sistemas de monitoreo y evaluación en cinco países latinoamericanos: Argentina, Chile, Colombia, Costa Rica y Uruguay*. Washington, DC: Banco Mundial.
- Zapico, E. (2005), “Desarrollo integrado de la evaluación y el presupuesto por resultados: un camino largo y pedregoso con destino incierto”, en Echebarría, K. *et al.*, *Responsabilización y evaluación de la gestión pública*. Caracas, Venezuela: CLAD-AECI/MAP/FIIAPP.

Otras referencias

- Acuerdo por el que se establece y da a conocer al público en general la denominación única de los acuíferos reconocidos en el territorio de los Estados Unidos Mexicanos, *Diario Oficial de la Federación*, 5 de diciembre de 2001.
- Acuña, S (s/f), “Diseño de evaluación al programa nacional de apoyo a la micro y pequeña empresa, fideicomiso 77-95 PRONAMYPE/BANCOOP RL”, tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- AEMA (2001a), *Publicaciones* [http://www.eea.europa.eu/es#tab-publications], fecha de consulta: 12 de octubre de 2014.
- (2001b), *Environmental signals. European environment agency regular indicator report* [www.eea.europa.eu/publications/signals-2001/signals2001], consultada por última vez el 23 de octubre de 2014.
- Alarcón, A. (s/f), “Evaluación de impacto de la oferta de capacitación del Centro de Imagen-INA (1995-1996)”, tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- American Evaluation Association [http://www.eval.org/], fecha de consulta: 4 de agosto de 2014.
- American Journal of Evaluation* [http://aje.sagepub.com/], fecha de consulta: 5 de agosto de 2014.
- Ardila, A. (2013), *Bases metodológicas para definir una agenda nacional de evaluación de política pública* [http://redlacme.org/profiles/blogs/bases-metodol-gicas-para-definir-una-agenda-nacional-de-evaluaci], fecha de consulta: 27 de febrero de 2013.
- Assemblée Nationale (2014), *Évaluations achevées et évaluation en cours* [www.assemblee-nationale.fr/14/controle/com_ccc/], fecha de consulta: 6 de diciembre de 2014.

- ASF (s/f), *Informe sobre la fiscalización superior de la política ambiental 2001-2008* [www.asf.gob.mx/trans/informes/IR2008i/Tomos/T8V2.pdf], fecha de consulta: 15 de agosto de 2012.
- (2010), *Informe del resultado de la fiscalización superior de la cuenta pública 2010* [http://www.asf.gob.mx/trans/Informes/IR2010i/indice.htm], fecha de consulta: 15 de agosto de 2012.
- (2011), *Informe del resultado de la fiscalización superior de la cuenta pública 2011. Evaluación de la política pública ambiental* [www.asf.gob.mx/Section/58_Informes_de_auditoria], fecha de consulta: 15 de diciembre de 2013.
- Banco Mundial (s/f), *México: análisis de los Lineamientos generales para la evaluación de los programas federales de la administración pública federal* [http://www.coneval.gob.mx/rw/resource/coneval/quienes/2937.pdf], fecha de consulta: 14 de mayo de 2013.
- (2009), *Agriculture and rural development public expenditure review* [http://siteresources.worldbank.org/INTMEXICO/Resources/EnglishPERDec16.pdf], fecha de consulta: 20 octubre de 2014.
- Baslé, M. *et al.* (2013), “Mobiliser la méta-évaluation pour mieux maîtriser la qualité des connaissances issues des évaluations dans la décision publique”, ponencia presentada en las *Jornadas Francesas de Evaluación*, SFE, Grenoble [http://hal.archives-ouvertes.fr/docs/00/86/38/36/PDF/bx2013-pub00038740.pdf], fecha de consulta: 10 de septiembre de 2014.
- Bourdin, J. *et al.* (2004), *Rapport d'information sur l'évaluation de politiques publiques en France* [www.senat.fr/rap/r03-392/r03-3921.pdf], fecha de consulta: 5 de diciembre de 2014.
- Burck, J. *et al.* (2011), *The climate change performance index, results 2012* [http://germanwatch.org/klima/ccpi.pdf], fecha de consulta: 12 de septiembre de 2014.
- Bustelo, M. (2001), *La evaluación de las políticas públicas de igualdad de género de los gobiernos central y autonómicos en España: 1995-1999*, tesis doctoral de la Universidad Complutense de Madrid, Madrid, España.
- (2013), *Documento temático sobre la credibilidad* [http://api.ning.com/files/.pdf], fecha de consulta: 15 de julio de 2014.
- C230 Consultores [http://www.c-230.com/], fecha de consulta: 19 de diciembre de 2014.
- (2012), *Evaluación de impacto no experimental del Programa Escuelas de Calidad 2012* [http://basica.sep.gob.mx/EVALUACION%20DE%20IMPACTO%20NO%20EXPERIMENTAL%20DEL%20PROGRAMA%20ESCUELAS%20DE%20CALIDAD.pdf], fecha de consulta: 19 de diciembre de 2014.
- Calvo, H. (s/f), “Evaluación de efectos de las iniciativas económicas innovadoras del Programa de Desarrollo Sostenible de Amigos de la Tierra España en las comunidades transfronterizas de Upala (Costa Rica) y el refugio de vida silvestre Los Guatusos

- (Nicaragua)", tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- Cámara de Diputados (2014, 14 de octubre), *Comisión de Desarrollo Social aprueba dar plena autonomía al Coneval*, Boletín de Comunicación Social, núm. 4379.
- (2014, 29 de octubre), *Avalan diputados dictamen que expide Ley del Coneval; garantizan plena autonomía*, Boletín de Comunicación Social núm. 4506.
- Canadian Evaluation Society [http://www.evaluationcanada.ca/], fecha de consulta: 12 de septiembre de 2014.
- Canadian Journal of Program Evaluation [http://www.evaluationcanada.ca/cjpe], fecha de consulta: 12 de septiembre de 2014.
- Cejudo, G. (Coord.) (2012), *Metaevaluación del Programa de Estancias Infantiles (PEI)* [http://www.2006-2012.sedesol.gob.mx/work/models/SEDESOL/EvaluacionProgramasSociales/2012/META_EV_2012/PEI_MetaEvaluacion.pdf], fecha de consulta: 12 de diciembre de 2014.
- (2014), *Implicaciones de la autonomía del Coneval* [http://rendiciondecuentas.org.mx/implicaciones-de-la-autonomia-constitucional-de-coneval/], fecha de consulta: 3 de septiembre de 2014.
- Centeno, H. (2002), "Meta-evaluación del Programa Antidrogas y Salud Sexual y Reproductiva para Adolescentes, Jamaica 1999-2002: lecciones aprendidas y contribución metodológica", tesis del Master Profesional en Población y Salud, Costa Rica: Universidad de Costa Rica.
- Centro CLEAR [http://www.clear-la.cide.edu/node/6], fecha de consulta: 15 de octubre de 2014.
- CMM (2012), *Evaluación del Programa de Acción Climática de la Ciudad de México 2008-2012*, México: GDF, Secretaría de Medio Ambiente/CMM.
- (2014), *Estrategia local de acción climática, Ciudad de México 2014-2020*, México: GDF, Sedema/CDMX/CMM.
- CEPAL (2011), *La planeación en Colombia: historia, actores e instrumentos* [https://sinergia.dnp.gov.co/Sinergia/Archivos/b795f9d4-783f-4ff8-b4d1-9020142387e9/ILPES_CEPAL_EL%20SALVADOR.pdf], fecha de consulta: 15 de agosto de 2014.
- CIDE (2014), *Las implicaciones de la autonomía del Coneval. Síntesis de temas y argumentos*. Foro de discusión del CIDE y otros organismos convocantes, mimeo.
- CLAD (2012), *Chile: principales logros y desafíos del sistema de evaluación y control de gestión* [http://siare.clad.org/siare/innotend/evaluacion/chile2/logros.html], fecha de consulta: 27 de noviembre de 2014.
- Collins, J. (2014), *Meta evaluation of local major schemes. Final report* [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/339175/meta-evaluation-final-report.pdf], fecha de consulta: 30 de agosto de 2014.

REFERENCIAS

- Commissariat Général du Plan [www.evaluation.gouv.fr/cgp], fecha de consulta: 9 de septiembre de 2014.
- Community Evaluation Solution [<http://www.communityevaluationsolutions.com>], fecha de consulta: 25 de septiembre de 2014.
- Conanp (s/f1), *Términos de referencia para la elaboración de programas de manejo de las áreas naturales protegidas competencia de la federación* [www.conanp.gob.mx/que_hacemos/pdf/programas_manejo/TERMINOS%20DE%20REF-PAGINA.pdf], fecha de consulta: 12 de septiembre de 2014.
- Conanp (2004), *Informe del sistema de información, monitoreo y evaluación para la conservación* [www.conanp.gob.mx/contenido/pdf/simec_2004.pdf], fecha de consulta: 18 de noviembre de 2012.
- Conanp (2006), *Informe del sistema de información, monitoreo y evaluación para la conservación* [www.conanp.gob.mx/contenido/pdf/simec_2006.pdf], fecha de consulta: 18 de noviembre de 2012.
- Conanp (2007), *Programa Nacional de Áreas Naturales Protegidas 2007-2012* [www.conanp.gob.mx/quienes_somos/pdf/programa_07012.pdf], fecha de consulta: 10 de diciembre de 2013.
- Conanp (2009), *Revisión y evaluación de medio término sobre la ejecución del Programa Nacional de Áreas Naturales Protegidas 2007-2012* [www.conanp.gob.mx/quienes_somos/pdf/PNANP2007-2012.pdf], fecha de consulta: 15 de abril de 2012.
- (2014), *Programa Nacional de Áreas Naturales Protegidas 2014-2018* [entorno. conanp.gob.mx/documento/pnanp.pdf], fecha de consulta: 15 de octubre de 2015.
- Conagua (2011), *Atlas del agua en México*. Distrito Federal: Semarnat.
- (2012), *Atlas del agua en México*. Distrito Federal: Semarnat.
- Coneval [<http://www.coneval.gob.mx/Paginas/principal.aspx>], fecha de consulta: 19 de diciembre de 2014.
- (s/f1), *Guion del análisis de factibilidad para llevar a cabo una evaluación de impacto* [http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/EVALUACIONES_PROGRAMAS_POLITICAS_DS/Evaluaciones%20de%20Impacto/guion_analisis_de_factibilidad.pdf], fecha de consulta: 3 de noviembre de 2014.
- (s/f2), *Modelo de términos de referencia para la evaluación en materia de diseño* [<http://www.coneval.gob.mx/Evaluacion/Paginas/Normatividad/TerminosReferencia2011.aspx>], fecha de consulta: 25 octubre de 2014.
- (s/f3), sin título (contiene información complementaria de la evaluación del Sistema de Protección Social en Salud, Seguro Popular) [http://www.coneval.gob.mx/rw/resource/coneval/eval_mon/1743.pdf], fecha de consulta: 23 de noviembre de 2014.

- (s/f4), *Evaluaciones integrales de la política social* [http://www.coneval.gob.mx/Evaluacion/Paginas/Evaluaciones_Programas/Evaluacion_Integral_Desempeno/Evaluacion_IDesempe%C3%B1o.aspx], fecha de consulta: 13 de diciembre de 2014.
- (s/f5), *Evaluaciones estratégicas de la política social* [http://www.coneval.gob.mx/Evaluacion/Paginas/Evaluaciones_Programas/Evaluacion_Estrategica/evaluacion_estrategica.aspx], fecha de consulta: 13 de diciembre de 2014.
- (2007a), *Modelo de términos de referencia para la evaluación en materia de diseño* [http://www.coneval.gob.mx/Evaluacion/Paginas/Evaluaciones_Programas/Evaluacion_Disenio/Evaluaciones_Disenio.aspx], fecha de consulta: 28 de octubre de 2014.
- (2007b), *Modelo de términos de referencia para la evaluación de consistencia y resultados* [http://www.coneval.gob.mx/rw/resource/coneval/eval_mon/906.pdf], fecha de consulta: 28 de octubre de 2014.
- (2009a), *Modelo de términos de referencia para la evaluación de procesos 2009* [http://www.coneval.gob.mx/Evaluacion/Paginas/Evaluaciones_Programas/Evaluacion_Procesos/Evaluacion_Procesos.aspx], fecha de consulta: 28 de octubre de 2014.
- (2009b), *Informe de seguimiento a los aspectos susceptibles de mejora de programas federales 2008* [http://www.coneval.gob.mx/rw/resource/coneval/eval_mon/2733.pdf], fecha de consulta: 30 de septiembre de 2014.
- (2010a), *Medición de la pobreza* [http://www.coneval.gob.mx/Informes/Interactivo/interactivo_nacional.swf], fecha de consulta: 3 de noviembre de 2014.
- (2010b), *Informe de la evaluación específica de desempeño 2009-2010. Programa Escuelas de Tiempo Completo* [http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/especificas_desempeno/dependencias/SEP/SEP34A.pdf], fecha de consulta: 29 de octubre de 2014.
- (2010c), *Informe de seguimiento a los aspectos susceptibles de mejora de programas federales 2009* [http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/Informe_Mecanismo_2009.pdf], fecha de consulta: 30 de septiembre de 2014.
- (2011a), *Elementos mínimos a considerar en la elaboración de diagnósticos de programas nuevos* [http://www.coneval.gob.mx/Informes/Evaluacion/Impacto/Diagnostico_Programas_Nuevos.pdf], fecha de consulta: 14 de octubre de 2014.
- (2011b), *Desempeño de los programas sociales* [www.coneval.gob.mx/cmsconeval/rw/pages/evaluacion/desempeno_programas_sociales.es.do], fecha de consulta: 10 de septiembre de 2012.
- (2011c), *Informe de la evaluación específica de desempeño 2010-2011* [www.coneval.gob.mx/Evaluacion/paginas/Fichas-narrativas-EED.aspx], fecha de consulta: 13 de junio de 2013.
- (2011d), *Valoración del desempeño de los programas* [<http://web.coneval.gob.mx/Evaluacion/Paginas/PaqueteInformativoPresupuestario.aspx>], fecha de consulta: 13 de junio de 2013.

- (2011e), *Seguimiento de aspectos susceptibles de mejora clasificados como específicos. Programa Escuelas de Tiempo Completo* [http://www.coneval.gob.mx/rw/resource/Mecanismos_2011/SEP/PETC_DT.pdf], fecha de consulta: 3 de diciembre de 2014.
- (2011f), *Seguimiento a aspectos susceptibles de mejora clasificados como específicos. Programa Escuelas de Calidad* [http://www.coneval.gob.mx/rw/resource/Mecanismos_2011/SEP/PEC_DT.pdf], fecha de consulta: 3 de diciembre de 2014.
- (2011g), *Informe de seguimiento a los aspectos susceptibles de mejora de programas federales 2010* [http://www.coneval.gob.mx/Evaluacion/Paginas/SR_Evaluaciones_E/2010-2011/Recomendaciones-de-Evaluacion-E-2010-2011.aspx], fecha de consulta: 30 de septiembre de 2014.
- (2011h), *Modelo de términos de referencia para la evaluación de consistencia y resultados* [http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/NORMATIVIDAD/TER_REF_2011/TDR_ECR.pdf], fecha de consulta: 28 de octubre de 2014.
- (2011i y 2012a), *Inventario de programas y acciones federales de desarrollo social, 2011/2012* [<http://www.coneval.gob.mx/evaluacion/ipfe/Paginas/default.aspx>], fecha de consulta: 14 de octubre de 2014.
- (2012b), *Ficha de desempeño del Programa Escuelas Dignas. 2012* [http://www.coneval.gob.mx/Informes/Evaluacion/Ficha_Monitoreo_2012/SEP/11_U074_FM.pdf], fecha de consulta: 7 de noviembre de 2014.
- (2012c), *Integración de ASM. Promanp 2012* [http://web.coneval.gob.mx/rw/resource/Mecanismos_2012/SEMARNAT/U035_Programa%20de%20Manejo%20de%20Areas%20Naturales%20Protegidas/U035_PI.pdf], fecha de consulta: 5 de marzo de 2014.
- (2012d), *Evaluación de consistencia y resultados 2011-2012. Programa de Escuelas de Tiempo Completo* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 3 de diciembre de 2014.
- (2012e), *Informe de seguimiento a los aspectos susceptibles de mejora de los programas y acciones federales de desarrollo social 2011-2012* [http://web.coneval.gob.mx/Informes/Evaluacion/Informe%20ASM/Informe%20ASM%20_2011_2012.pdf], fecha de consulta: 30 de septiembre de 2014.
- (2012f), “Evaluación de políticas públicas de desarrollo social en México”, ponencia presentada en el *III Encuentro de la Red colombiana de monitoreo y evaluación de políticas públicas*, Bogotá, Colombia.
- (2012g), *Medición del índice de difusión de las evaluaciones externas de los programas federales de desarrollo social 2012* [http://www.coneval.gob.mx/Informes/Evaluacion/Impacto/Analisis_Difusion/INDICADOR_DIFUSION_2012.pdf], fecha de consulta: 22 de septiembre de 2014.

- (2013a), *Modelo de términos de referencia para la evaluación específica de desempeño 2012-2013* [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=xd9kVMj6OZWhugS6pYDoAw&ved=0CAkQFjADOB4&client=internal-uds-cse&usg=AFQjCNFWz_K8tp0j-3hwDib-7B3AQ4R3nA], fecha de consulta: 12 de noviembre de 2014.
- (2013b), *Programa de Desarrollo Humano Oportunidades. Ficha de monitoreo 2013* [http://www.coneval.gob.mx/Informes/Evaluacion/Ficha_Monitoreo_Evaluacion_2013/SEDESOL/20_S072.pdf], fecha de consulta: 12 de noviembre de 2014.
- (2013c), *Programa de Escuelas de Tiempo Completo. Ficha de monitoreo 2013* [http://www.coneval.gob.mx/Informes/Evaluacion/Ficha_Monitoreo_Evaluacion_2013/SEP/11_S221.pdf], fecha de consulta: 12 de noviembre de 2014.
- (2013d), *Ficha de desempeño del Programa Escuelas Dignas, 2013* [http://www.coneval.gob.mx/Informes/Evaluacion/Ficha_Monitoreo_Evaluacion_2013/SEP/11_U074.pdf], fecha de consulta: 7 de noviembre de 2014.
- (2013e), *Programa anual de evaluación para el ejercicio fiscal 2013 de los programas de la administración pública federal* [http://www.coneval.gob.mx/Informes/Normateca/PAE_2013.pdf], fecha de consulta: 16 de octubre de 2014.
- (2013f), *Informe de la evaluación específica de desempeño 2012 - 2013. Programa de Apoyo al Ingreso Agropecuario: Procampo para Vivir Mejor; Valoración de la información de desempeño presentada por el programa* [<http://www.sagarpa.gob.mx/programas2/evaluacionesExternas/Documents/Evaluaciones%20Espec%20C3%ADficas%20de%20Desempe%20C3%B1o%20%28EED%29%202012%20-%202013/S231%20Ejecutivo.pdf>], fecha de consulta: 16 de octubre de 2014.
- (2013g), *Evaluación estratégica de protección social en México* [http://www.coneval.gob.mx/Informes/Evaluacion/Estrategicas/Evaluacion_Estrategica_de_Proteccion_Social_en_Mexico.pdf], fecha de consulta: 5 de noviembre de 2014.
- (2013h), *Informe de la evaluación específica de desempeño 2012-2013. Programa de Desarrollo Humano Oportunidades* [http://www.coneval.gob.mx/Informes/Evaluacion/Especificas_Desempeno2012/SEDESOL/20_S072/20_S072_Ejecutivo.pdf], fecha de consulta: 7 de octubre de 2014.
- (2013i), *Modelo de términos de referencia para la evaluación de procesos del programa (nombre del programa)* [http://www.coneval.gob.mx/rw/resource/coneval/EVALUACIONES/EVALUACION_DE_PROCESOS.pdf], fecha de consulta: 10 de noviembre de 2014.
- (2013j), *Informe de seguimiento a los aspectos susceptibles de mejora de los programas y acciones federales de desarrollo social 2012-2013* [<http://www.coneval.gob.mx/Informes/Evaluacion/Informe%20ASM/Informe%20ASM%202012-2013.pdf>], fecha de consulta: 30 de septiembre de 2014.

- (2013k), *Seguimiento a aspectos susceptibles de mejora clasificados como específicos, derivados de informes y evaluaciones externas. Programa de Desarrollo Humano Oportunidades* [http://www.coneval.gob.mx/rw/resource/Mecanismos_2012/SEDESOL/S072_Programa%20de%20Desarrollo%20Humano%20Oportunidades/S072_DT.pdf], fecha de consulta: 7 de octubre de 2014.
- (2013l), *Seguimiento a aspectos susceptibles de mejora clasificados como institucionales, derivados e informes y evaluaciones externas* [http://www.coneval.gob.mx/rw/resource/Mecanismos_2012/SEDESOL/SEDESOL_DI.pdf], fecha de consulta: 7 de octubre de 2014.
- (2013m), *Fichas de monitoreo de programas de desarrollo social 2012-2013* [http://www.coneval.gob.mx/Evaluacion/Eval_Especificas/Paginas/FM_2012-2013/Ficha_Monitoreo_2012.aspx], fecha de consulta: 8 de octubre de 2014.
- (2014a), *Realiza el Coneval segundo seminario para entidades federativas* [<http://www.coneval.gob.mx/Eventos/Paginas/Eventos.aspx>], fecha de consulta: 15 de noviembre de 2014.
- (2014b), *Informe de seguimiento a los aspectos susceptibles de mejora de los programas y acciones federales de desarrollo social 2013-2014* [<http://www.coneval.gob.mx/Informes/Evaluacion/Informe%20ASM/Informe%20ASM%202013-2014.pdf>], fecha de consulta: 30 de septiembre de 2014.
- (2014c), *Seguimiento a aspectos susceptibles de mejora clasificados como específicos, derivados de informes y evaluaciones externas. Estabilización de Cuencas y Acuíferos* [http://www.semarnat.gob.mx/sites/default/files/documentos/programas/evaluaciones/asm/2014/abril/k141_dt.pdf], fecha de consulta: 14 de enero de 2015.
- (2014d), *Autonomía de hecho y autonomía de derecho: los retos de la reforma política para el Coneval* [http://www.coneval.gob.mx/Informes/COMUNICADOS_DE_PRENSA/REFLEXIONES_SOBRE_LA_AUTONOMIA_DEL_CONEVAL.pdf], fecha de consulta: 3 de septiembre de 2014.
- Coneval (varios años), *Buenas prácticas en monitoreo y evaluación* [www.coneval.gob.mx/Evaluacion/Paginas/Buenas_Practicas/BP_2009;BP_2010;BP_2011;BP_2013-2014] [http://www.coneval.gob.mx/SalaPrensa/Paginas/Comunicados_de_prensa/Buenas-Pr%C3%A1cticas--2012.aspx], fechas de consulta: 30 de mayo de 2013 y 20 de octubre de 2014.
- Coneval/SHCP/SFP (2007), *Lineamientos generales para la evaluación de los programas federales de la administración pública federal* [http://www.coneval.gob.mx/rw/resource/coneval/eval_mon/361.pdf], fecha de consulta: 3 de septiembre de 2014.
- Congressional Budget Office [<https://www.cbo.gov/>], fecha de consulta: 15 de junio de 2014.
- Consejo Centroamericano de Acreditación de la Evaluación Superior *Revista CCAIidad* [http://www.cca.ucr.ac.cr/publicaciones/revista_ccalidad], fecha de consulta: 3 de octubre de 2014.

- Decreto de Presupuesto de Egresos de la Federación 2013, *Diario Oficial de la Federación*, 27 de diciembre de 2012.
- Decreto por el que se crea el Instituto Nacional para la Evaluación de la Educación, *Diario Oficial de la Federación*, 8 de agosto de 2002.
- Decreto por el que se reforma el diverso por el que se crea el Instituto Nacional para la Evaluación de la Educación, publicado el 8 de agosto de 2012, *Diario Oficial de la Federación*, 16 de mayo de 2012.
- Decreto por el que se reforman los artículos 3 en sus fracciones III, VII y VIII; y 73, fracción XXV y se adiciona un párrafo tercero, un inciso d) al párrafo segundo de la fracción II y una fracción IX al artículo 3 de la Constitución Política de los Estados Unidos Mexicanos, *Diario oficial de la Federación*, 26 de febrero de 2013.
- De Faria, H.H. (1993), *Elaboración de un procedimiento para medir la efectividad de manejo de áreas silvestres protegidas y su aplicación en dos áreas protegidas de Costa Rica*. Tesis magister Scientiae. Turrialba, Costa Rica, CATIE.
- De la Garza, T. (2010), “Evaluación de la política y los programas de desarrollo social”, ponencia presentada en el CIDE, México.
- De la Maza, E.J. et al. (2003), *Estado actual de las áreas naturales protegidas de América Latina y el Caribe*. PNUMA/Quercus Consultoría Ecológica, SC.
- Díaz, L.F. (2005), “Metaevaluación del modelo del sistema nacional de evaluación de Costa Rica”, ponencia presentada en el *X Congreso Internacional del CLAD*, Santiago de Chile.
- Ministerio de Hacienda, Dipres (s/f), *Sistema de evaluación y control* [<http://www.dipres.gob.cl/594/w3-propertyvalue-2131.html>], fecha de consulta: 30 de noviembre de 2014.
- Dipres (s/f): *Sistema de evaluación y control* [<http://www.dipres.gob.cl/594/w3-propertyvalue-2131.html>], fecha de consulta: 30 de noviembre de 2014.
- Dipres (2008), *Notas técnicas* [http://www.dipres.cl/control_gestion/publicaciones/Notas_tecnicas_enero2008.html], fecha de consulta: 25 de noviembre de 2014.
- (2012), *Programa de Mejoramiento de Gestión* [<http://www.dipres.gob.cl/594/w3-propertyvalue-15230.html>], fecha de consulta: 25 de noviembre de 2014.
- DNP [<https://www.dnp.gov.co>], fecha de consulta: 19 de diciembre de 2014.
- (2011a), *Evaluación del desempeño integral de los municipios* [<https://www.dnp.gov.co/programas/desarrollo-territorial/evaluacion-y-seguimiento-de-la-descentralizacion/Paginas/documentos-de-evaluacion.aspx>], fecha de consulta: 1 de septiembre de 2014.
- (2011b), *La planeación en Colombia: historia, actores e instrumentos* [https://sinergia.dnp.gov.co/Sinergia/Archivos/b795f9d4-783f-4ff8-b4d1-9020142387e9/ILPES_CEPAL_EL%20SALVADOR.pdf], fecha de consulta: 15 de agosto de 2014.
- Embajada de Francia en México (2014) [<http://www.ambafrance-mx.org/La-Agencia-Francesa-de-Desarrollo>], fecha de consulta: 12 de septiembre de 2014.

- Escolero, O. (2013), *Evaluación del diseño del programa Estabilización de Cuencas y Acuíferos* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 19 de diciembre de 2014.
- European Evaluation Society [<http://www.europeanevaluation.org/>], fecha de consulta: 12 de octubre de 2014.
- EvalPartners Innovation Challenge (2013), *Evaluations that make a difference* [<https://evaluationstories.wordpress.com/>], fecha de consulta: 24 de octubre de 2014.
- EvalSolutions [<http://www.evulsolutions.com/Esportal>], fecha de consulta: 25 de septiembre de 2014.
- Evalúa DF (s/f1), Documentos diversos (respuesta de los organismos públicos a las evaluaciones de 2010 y a sus recomendaciones) [http://www.evalua.df.gob.mx/evaluacion-recomendaciones_10.php], fecha de consulta: 18 de febrero de 2014.
- (s/f2), *Recomendaciones a la política de equidad de género del gobierno del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/recomendaciones_comite/rec_equidad_genero.pdf], fecha de consulta: 28 de agosto de 2014.
- (s/f3), *Recomendaciones a la política de participación ciudadana en el desarrollo social del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/recomendaciones_comite/reco_parti_ciuda.pdf], fecha de consulta: 5 de septiembre de 2014.
- (s/f4), *Recomendaciones a la política de transporte colectivo de pasajeros del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/comentarios/reco_transporte.pdf], fecha de consulta: 23 de septiembre de 2014.
- (2009a) Estatuto Orgánico, *Gaceta Oficial del Distrito Federal*, 15 de junio de 2009.
- (2009b), *Recomendaciones al Programa de Estímulos para el Bachillerato Universal (Prepa Si)* [www.evalua.df.gob.mx/recomendaciones/recomendaciones_comite/rec_def_prepa.pdf], fecha de consulta: 19 de julio de 2012.
- (2010a), *Convocatoria 001/2010 para la evaluación externa de las políticas de desarrollo social del Distrito Federal* [http://www.evalua.df.gob.mx/files/convocatoria/2010/conv_2010.pdf], fecha de consulta: 15 de agosto de 2014.
- (2010b), *Convocatoria 002/2010 para la evaluación externa de la política de equidad de género en el Distrito Federal* [http://www.evalua.df.gob.mx/files/convocatoria/2010/con_2_10.pdf], fecha de consulta: 15 de agosto de 2014.
- (2010c), *Términos de referencia para la evaluación del diseño e instrumentación de la política de equidad de género del gobierno del Distrito Federal* [http://www.evalua.df.gob.mx/files/convocatoria/2010/tr_gener_10.pdf], fecha de consulta: 15 de agosto de 2014.
- (2010d), *Términos de referencia para la evaluación del diseño e instrumentación de la política de transporte público colectivo de pasajeros en el Distrito Federal* [http://www.evalua.df.gob.mx/files/convocatoria/2010/tr_transp_10.pdf], fecha de consulta: 15 de agosto de 2014.

- (2010e), *Términos de referencia para la evaluación del diseño, instrumentación y resultados de la política de participación ciudadana en el desarrollo social del gobierno del Distrito Federal* [http://www.evalua.df.gob.mx/files/convocatoria/2010/tr_participa_10.pdf], fecha de consulta: 15 de agosto de 2014.
- (2010f), *Propuesta metodológica: evaluación del diseño, instrumentación y resultados de la política de participación ciudadana en el desarrollo social del gobierno del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/evaluaciones_finales/prop_met.pdf], fecha de consulta: 5 de septiembre de 2014.
- (2010g), *Recomendaciones al Programa Pensión Alimentaria para Adultos Mayores de 70 años, residentes en el Distrito Federal* [www.evalua.df.gob.mx/recomendaciones/recomendaciones_comite/rec_def_pen_al.pdf], fecha de consulta: 19 de julio de 2012.
- (2010h), *Recomendaciones a la política de acceso al agua potable* [http://www.evalua.df.gob.mx/recomendaciones/recomendaciones_comite/rec_agua.pdf], fecha de consulta: 19 de julio de 2012.
- (2010i), *Recomendaciones a las políticas de fomento al empleo y protección del trabajo* [http://www.evalua.df.gob.mx/transparencia/2010/rec_tra.pdf], fecha de consulta: 19 de julio de 2012.
- (2011a), *Informe final de evaluación de la política de equidad de género en el gobierno del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/evaluaciones_finales/eva_eq_gen.pdf], fecha de consulta: 25 de agosto de 2014.
- (2011b), *Informe final de la evaluación del diseño, instrumentación y resultados de la política de participación ciudadana en el desarrollo social del Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/evaluaciones_finales/ev_part_ciu.pdf], fecha de consulta: 2 de septiembre de 2014.
- (2011c), *Entrega final de la evaluación del diseño e instrumentación de la política de transporte público colectivo de pasajeros en el Distrito Federal* [http://www.evalua.df.gob.mx/files/recomendaciones/evaluaciones_finales/ev_transp.pdf], fecha de consulta: 19 de septiembre de 2014.
- (2011d), *Convocatoria 001/2011 para la evaluación externa de políticas y programas sociales del GDF* [http://www.evalua.df.gob.mx/files/convocatoria/2011/conv_11_eva.pdf], fecha de consulta: 15 de agosto de 2014.
- (2011e), *Términos de referencia para la evaluación de operación, resultados e impacto de los servicios de salud materno-infantil del gobierno del Distrito Federal 2006-2010* [http://www.evalua.df.gob.mx/files/convocatoria/2011/ter_sal_11.pdf], fecha de consulta: 15 de agosto de 2014.
- (2011f), *Informe final de la evaluación de operación, resultados e impactos de los servicios de salud materno-infantil del gobierno del Distrito Federal 2006-2010* [<http://www2.df.gob>].

- mx/virtual/evaluadf/files/recomendaciones/evaluaciones_finales/info_final.pdf], fecha de consulta: 6 de octubre de 2014.
- (2011g), *Primer informe semestral del estado de las recomendaciones aceptadas*, mimeo. Evaluation Consultation Services [<http://www.diehlconsulting.org>], fecha de consulta: 25 de septiembre de 2014.
- Evaluation Review [<http://erx.sagepub.com/>], fecha de consulta: 5 de agosto de 2014.
- Evaluation and Program Planning Journal [<http://www.journals.elsevier.com/evaluation-and-program-planning>], fecha de consulta: 4 de agosto de 2014.
- FAO (2006), *Sistematización de la experiencia del proyecto evaluación de Alianza para el Campo* [<http://www.fao-evaluacion.org.mx/pagina/documentos/estudios-especiales/Sistematizacion-de-la-experiencia-del-proyecto.pdf>], fecha de consulta: 30 de noviembre de 2014.
- (2013), *Evaluación nacional del funcionamiento y la operación 2012 del Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural. Componentes en concurrencia de recursos* [http://www.fao-evaluacion.org.mx/evaluacion_2012/informes/pdcityer_web/files/pdciter_int.pdf], fecha de consulta: 20 de octubre de 2014.
- Fatterman & Associates [<http://www.davidfetterman.com>], fecha de consulta: 25 de septiembre de 2014.
- Foro Sistema de Evaluación del Desempeño (en gobiernos locales) para Presupuesto basado en Resultados (2014, 20 de noviembre) [http://redlacme.org/profiles/blog/show?id=4315104:BlogPost:95534&xgs=1&xg_source=msg_share_post], fecha de consulta: 29 de octubre de 2014.
- Fox, J. y L. Haight (2010), *Evaluating evaluators. Subsidizing inequality: mexican corn policy since NAFTA*. Woodrow Wilson International Center for Scholars [http://www.wilsoncenter.org/sites/default/files/Subsidios_Cap_1_Fox%20and%20Haight.pdf], fecha de consulta: 6 de octubre de 2013.
- Franco, R. (2003), “Descentralización, participación y competencia en la gestión social”, ponencia presentada en el *VIII Congreso Internacional del CLAD*, Panamá.
- Fulcrum Corporation [<http://www.fulcrum-corp.com>], fecha de consulta: 25 de septiembre de 2014.
- Gakidou, E. *et al.* (2005), *Evaluación del impacto de la reforma mexicana de salud 2001-2006: un informe inicial* [<http://bvs.insp.mx/rsp/articulos/articulo.php?id=001802>], fecha de consulta: 2 de diciembre de 2014.
- GAO [<http://www.gao.gov>], fecha de consulta: 10 de julio y 15 de agosto de 2014.
- García, E. (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.
- Gargani + Company [<http://gcoinc.com>], fecha de consulta: 25 de septiembre de 2014.
- GDF (2007), *Programa de Desarrollo Social 2007-2012 de la Ciudad de México* [<http://cgsservicios.df.gob.mx/prontuario/vigente/1695.htm>], fecha de consulta: 11 de septiembre de 2014.
- (2011), *Informe de avances del Programa de Acción Climática de la Ciudad de México 2008-2012*, GDF, México.
- Gesoc (2014), *Índice de desempeño de los programas públicos federales 2014* [<http://www.indep.gesoc.org.mx/>], fecha de consulta: 12 de noviembre de 2014.
- GIZ (2012a), *GIZ's Integrated results model: A working aid for appraisal and offer preparation*, Eschborn, Alemania.
- (2012b), *GIZ's Results-based monitoring system: framework of reference*, Eschborn, Alemania.
- Government Executive [<http://www.govexec.com/>], fecha de consulta: 19 de septiembre de 2014.
- Government of Canada. Canadian Heritage [<http://www.pch.gc.ca/eng/1343398573651/1343399673631>], fecha de consulta: 2 de diciembre de 2014.
- Guzmán, M. (2009), “Desafíos para la toma de decisiones basada en la evidencia. La experiencia chilena. Principales características de su diseño, implementación y uso de la información”, ponencia presentada en el *Seminario de monitoreo y evaluación*, Bogotá, Colombia.
- Hernández, T. (en proceso), *El uso instrumental de las evaluaciones en México. Estudio de caso: programa Caravanas de la Salud*, tesis de maestría en Políticas Públicas de la Universidad Autónoma Metropolitana-Xochimilco, México.
- Hernández, S. (2014), *Informe de labores 2012-2014. Mideplan* [<http://documentos.mideplan.go.cr:8080/alfresco/d/d/workspace/SpacesStore/9431d2b3-a41e-4b23-a8f6-0ce59d1f303f/Informe%20Viceministra%20Silvia.pdf>], fecha de consulta: 15 de agosto de 2014.
- HM Treasury [<https://www.gov.uk/government/organisations/hm-treasury>], fecha de consulta: 17 de julio de 2014.
- (2011), *The green book. Appraisal and evaluation in central government* [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf], fecha de consulta: 21 de julio de 2014.
- (2011), *The magenta book. Guidance for evaluation* [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf], fecha de consulta: 21 de julio de 2014.

- IDS (2008), *Evaluating adaptation to climate change from a development perspective* [http://www.unisdr.org/files/7845_GEF20final20report20Oct20081.pdf], fecha de consulta: 20 de octubre de 2014.
- IMCO (2013), *Evaluación del Programa Especial del Cambio Climático* [http://imco.org.mx/wp-content/uploads/2013/2/studie_2_pecc_web_ok4.pdf], fecha de consulta: 20 de agosto de 2014.
- INECC (2014), *Avances de los programas estatales de acción ante el cambio climático* [<http://www2.inecc.gob.mx/sistemas/peacc/>], fecha de consulta: 12 de septiembre de 2014.
- INEE [<http://www.inee.edu.mx/>], fecha de consulta: 10 de octubre de 2014.
- (2013), Estatuto Orgánico, *Diario Oficial de la Federación*, 22 de octubre de 2013.
- INEGI (2005), *Población rural y rural ampliada en México, 2000* [http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2000/pob_rural/Rural_y_Rural_Ampliada.pdf], fecha de consulta: 8 de diciembre de 2014.
- INGC (2009), *Study on the impact of climate change on disaster risk in Mozambique: synthesis report* [http://www.irinnews.org/pdf/synthesis_report_final_draft_march09.pdf], fecha de consulta: 11 de septiembre de 2014.
- Inifed (2014), *Programa institucional del Instituto Nacional de la Infraestructura Física Educativa* [http://dof.gob.mx/nota_detalle.php?codigo=5343877&fecha=08/05/2014], fecha de consulta: 7 de noviembre de 2014.
- Insad (2008), *Evaluación de efectos del Sistema de protección social en salud. Seguro Popular de Salud* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 17 de noviembre de 2014.
- INSP (2011), *Evaluación de impacto del Programa Estancias Infantiles para Apoyar a Madres Trabajadoras* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 4 de noviembre de 2014.
- International Organization Development (2012), *Meta evaluation of mine action and development. Final report* [<http://www.iodparc.com/documents/metaevaluation.pdf>], fecha de consulta: 15 de agosto de 2014.
- Irarrázaval, I. e I. Arteaga (2012), “Impacto de la evaluación: aprendizajes del caso chileno”, ponencia presentada en el *XVII Congreso Internacional del CLAD sobre la Reforma del Estado y la Administración Pública*, Cartagena, Colombia.
- Jaime, E. (2014), “Preocupa Coneval”, *Excélsior*, 1 de noviembre [<http://www.excelsior.com.mx/opinion/edna-jaime/2014/11/01/990040>], fecha de consulta: 12 de noviembre de 2014.
- JBS International Burlingame [<http://www.jbsinternational.com>], fecha de consulta: 25 de septiembre de 2014.
- Joint Committee on Standards for Educational Evaluation [www.jcsee.org], fecha de consulta: 18 de junio de 2014.

- Journal of MultiDisciplinary Evaluation [http://journals.sfu.ca/jmde/index.php/jmde_1/index], fecha de consulta: 4 de agosto de 2014.
- Kingsbury, N. (2005), *La evaluación de programas en Estados Unidos: historia y naturaleza del papel desempeñado por la oficina de supervisión del gobierno* [[http://www2.congreso.gob.pe/sicr/cendocbib/con3_uibd.nsf/11F49FC077D22EA3052578CC006F563B/\\$FILE/5_Eva.Programa.pdf](http://www2.congreso.gob.pe/sicr/cendocbib/con3_uibd.nsf/11F49FC077D22EA3052578CC006F563B/$FILE/5_Eva.Programa.pdf)], fecha de consulta: 2 de julio de 2014.
- Lahey, R. (2010), *The Canadian M&E System: lessons learned from 30 years of development*, Evaluation Capacity Development Working Paper Series, núm. 23. Washington, DC: The World Bank [http://siteresources.worldbank.org/INTEVACAPDEV/Resources/ecd_wp_23.pdf], fecha de consulta: 20 de octubre de 2014.
- Lajouidie, B. y J. Toulemonde (2002), “Méta-évaluation, bilan des évaluations réalisées en Limousin au cours de la période 1994-1998”, ponencia presentada en las *Cuarta Jornadas de la Société Française d'Évaluation*, Lille, Francia.
- Larrú, J.M. (2003), “Análisis de los resultados y metaevaluación del sistema de evaluaciones de EuropeAid”, España: CECOD [www.cecod.net], fecha de consulta: 22 de junio de 2004.
- Läubli, M. (2013), *UKES Annual conference notes 2013* [<http://www.seval.ch/documents/ukes2013.pdf>], fecha de consulta: 5 de diciembre de 2014.
- Ley de Desarrollo Social para el Distrito Federal, *Gaceta Oficial del Distrito Federal*, 23 de mayo de 2000.
- Ley del Instituto Nacional para la Evaluación de la Educación, *Diario Oficial de la Federación*, 11 de septiembre de 2013.
- Ley Federal de Presupuesto y Responsabilidad Hacendaria, *Diario Oficial de la Federación*, 30 de marzo de 2006 (reformada el 31 de diciembre de 2008).
- Ley General de Cambio Climático, *Diario Oficial de la Federación*, 6 de junio de 2012 (última reforma el 29 de diciembre de 2014).
- Ley General de Desarrollo Social, *Diario Oficial de la Federación*, 20 de enero de 2004 (última reforma 07 de noviembre de 2013).
- Ley General de Educación, *Diario Oficial de la Federación*, 13 de julio de 1993 (última reforma publicada el 20 de mayo de 2014).
- Ley General del Equilibrio Ecológico y la Protección al Ambiente, *Diario Oficial de la Federación*, 28 de enero de 1988 (última reforma publicada el 16 de enero de 2014).
- Library of Congress, Congressional Research Service [<http://www.loc.gov/crsinfo/about/>], fecha de consulta: 23 de septiembre de 2014.
- Mackay, K. (2006), *Institucionalización de los sistemas de seguimiento y evaluación para mejorar la gestión del sector público*, Evaluation Capacity Development Working Paper Series núm. 15, Washington, DC: Banco Mundial [<http://www-wds.worldbank.org/external/>]

- default/WDSContentServer/WDSP/IB/2006/12/29/000310607_20061229131947/Rendered/PDF/378280SPANISH01aluation1psm1spanish.pdf], fecha de consulta: 13 de octubre de 2014.
- Mahecha, J. (2007), “Una experiencia en metaevaluación de programas”, tesis de mágister en Educación Universidad de Los Andes, Bogotá, Colombia.
- Márquez, F. (2013), “Metaevaluación. Una mirada al sistema de evaluación de programas federales en México: El caso de tres evaluaciones externas al Programa de Desarrollo Humano Oportunidades”, tesis de maestría en políticas públicas, UAM-Xochimilco, México.
- Mayne, J. (2008), *Contribution analysis: an approach to exploring cause and effects* [http://www.bond.org.uk/data/files/Contribution_analysis__An_approach_to_exploring_cause_and_effect.pdf], fecha de consulta: 12 de diciembre de 2014.
- Medina, O. (2014), “Metaevaluación de los programas de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (Sagarpa)”, tesis de maestría en gobierno y asuntos públicos de la Universidad Nacional Autónoma de México, México.
- Merino, M. (2009), *Los programas de subsidios al campo: las razones y las sinrazones de una política mal diseñada* [http://www.wilsoncenter.org/sites/default/files/Subsidios_Cap_2_Merino.pdf], fecha de consulta: 6 de marzo de 2014.
- Mideplan (s/f), *Informes de evaluaciones y monitoreo de metas* [<http://www.mideplan.go.cr/el-plan-nacional-de-desarrollo/35-evaluacion/336-informacion-general.html>], fecha de consulta: 7 de octubre de 2014.
- Mideplan (2011), *Revisión experta al personal de Mideplan en metodologías para seguimiento y evaluación del PND y conformación del registro nacional de evaluaciones estratégicas del gobierno (2010-2011)* [<http://www.mideplan.go.cr/rendicion-cuentas/consultorias?id=1099>], fecha de consulta: 27 de octubre de 2014.
- Mideplan (2012), *Manual gerencial para el diseño y ejecución de evaluaciones estratégicas de gobierno*, Costa Rica.
- Mideplan (2014), *Informe anual de la matriz anual de programación y seguimiento sectorial e institucional* [<http://documentos.mideplan.go.cr:8080/alfresco/d/d/workspace/SpacesStore/58bd6557-e740-436c-a642-88e2430b1181/Informe%20MAPI%202013.pdf>], fecha de consulta: 15 de agosto de 2014.
- Mideplan-Foceval (2012), *Fortalecimiento de las capacidades en evaluación en Centroamérica 2011-2012* [<http://www.mideplan.go.cr/rendicion-cuentas/consultorias?id=1098>], fecha de consulta: 27 de octubre de 2014.
- Mideplan-Ministerio de Hacienda (2014), *Lineamientos técnicos y metodológicos para la planificación, programación presupuestaria y evaluación estratégica en el sector público en Costa Rica* [<http://documentos.mideplan.go.cr:8080/alfresco/d/d/workspace/SpacesStore/488e94b8->

- 117e-44ad-9ba0-27094beb20c0/Lineamientos_POI_2014%2016-05-2013.pdf], fecha de consulta: 1 de septiembre de 2014.
- Ministerio de Economía y Hacienda del Gobierno de España y Banco Mundial (s/f), *La formulación de políticas en la OCDE: ideas para América Latina* [http://siteresources.worldbank.org/INTLAC/Resources/OECD_IDEAS_spanish.pdf?cid=EXT_BoletinES_W_EXT], fecha de consulta: 18 de junio de 2014.
- Montiel, M. J. (2012), *¿Qué pasa con las evaluaciones a los programas de salud en México?* [<http://estepais.com/site/2012/%C2%BFque-pasa-con-las-evaluaciones-a-los-programas-de-salud-en-mexico/>], fecha de consulta: 23 de noviembre de 2014.
- Morales, C. (2009), “Arreglos institucionales para los sistemas de M&E en América Latina: experiencia de Costa Rica”, ponencia en la *V Conferencia de la Red de Monitoreo y Evaluación en América Latina y el Caribe*, Bogotá, Colombia.
- Morillo, L. (s/f), “Diseño de evaluación Programa Nacional de Atención Integral al Menor Trabajador y de la Calle Instituto Nacional de Bienestar Familiar INABIF-Lima, Perú”, tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- Morin, E. y J.L. Le Moigne (1999), “Inteligencia de la complejidad”, ponencia presentada en el *Coloquio de Cerisy*, Francia.
- Mosaic [<http://www.mosaic-net-intl.ca/whatis.shtml>], fecha de consulta: 8 de septiembre de 2014.
- NAO [<http://www.nao.org.uk/>], fecha de consulta: 8 de septiembre de 2014.
- Neirotti, N. (s/f), *La función de evaluación de programas sociales en Chile, Brasil y Argentina* [http://www.academia.edu/772262/La_funcion_de_evaluacion_de_programas_sociales_en_Chile_Brasil_y_Argentina], fecha de consulta: 19 de septiembre de 2014.
- New Directions for Evaluation [[http://onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1534-875X](http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1534-875X)], fecha de consulta: 4 de agosto de 2014.
- Núñez, J. (2012), *Evaluación de diseño del Programa Programas de Manejo de Áreas Naturales Protegidas (Promanp)* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 27 de octubre de 2014.
- OCDE (2013), *Evaluación de la OCDE sobre el desempeño ambiental: México*. México: OECD Publishing.
- Office for Budget Responsibility [<http://budgetresponsibility.org.uk/>], fecha de consulta: 14 de julio de 2014.
- Office of the Deputy Prime Minister (2005), *Meta-evaluation of the local government modernisation agenda: progress report in service improvement in local government* [<http://www.bipsolutions.com/docstore/pdf/9882.pdf>], fecha de consulta: 14 de julio de 2014.

REFERENCIAS

- Office of Management and Budget [<http://www.whitehouse.gov/omb>], fecha de consulta: 20 de junio de 2014.
- OMB (2010) [www.govexec.com/oversight/2010/08/omb-will-provide-funding-for-program-evaluations/32059/], fecha de consulta: 19 de septiembre de 2014.
- Ospina, S. y N. Cunill (2008), *Origen y antecedentes del sistema de evaluación y control de gestión* [<http://siare.clad.org/siare/innotend/evaluacion/chile2/todo.html>], fecha de consulta: 9 de mayo de 2014.
- Osuna, J.L. *et al.* (2011), *Metaevaluación de las políticas evaluadas por Evalúa DF durante 2009* [http://www.evalua.df.gob.mx/files/transparencia/2011/inf_finmet_09.pdf], fecha de consulta: 7 de marzo de 2017.
- PDHO (2008), *Resumen ejecutivo de la evaluación de consistencia y resultados 2007 del Programa de Desarrollo Humano Oportunidades* [http://www.oportunidades.gob.mx/EVALUACION/441c7c1a3d30adf64e0e724174a9d527/impacto/2007/ECR_2007_2008_ResumenEjecutivo.pdf], fecha de consulta: 8 de septiembre de 2012.
- (2009), *Informe de la evaluación específica de desempeño 2008 del Programa de Desarrollo Humano Oportunidades* [http://www.oportunidades.gob.mx/EVALUACION/441c7c1a3d30adf64e0e724174a9d527/eval_especifica/oportunidades/oport_informe_completo_2008_2009.pdf], fecha de consulta: 8 de septiembre de 2012.
- (2010), *Informe de la evaluación específica de desempeño 2009-2010 del Programa de Desarrollo Humano Oportunidades* [http://www.oportunidades.gob.mx/EVALUACION/441c7c1a3d30adf64e0e724174a9d527/eval_especifica/oportunidades/oport_informe_completo_2009_2010.pdf], fecha de consulta: 8 de septiembre de 2012.
- (2011), *Informe de la evaluación específica de desempeño 2010-2011 del Programa de Desarrollo Humano Oportunidades* [http://www.oportunidades.gob.mx/EVALUACION/441c7c1a3d30adf64e0e724174a9d527/eval_especifica/oportunidades/oport_informe_completo_2010_2011.pdf], fecha de consulta: 8 de septiembre de 2012.
- PECC 2009-2012 *Diario Oficial de la Federación*, 28 de agosto de 2009.
- Pnuma (s/f) [www.pnuma.org/deat1/pdf/Estado%20actual%20de%20las%20Areas%20Naturales%20Protegidas%20de%20America%20Latina%20y%20el%20Caribe.pdf], fecha de consulta: 11 de noviembre de 2012.
- Programa para un Gobierno Cercano y Moderno 2013-2018, *Diario Oficial de la Federación*, 30 de agosto de 2013.
- Quintero, V. (2010), *Características de las demandas de servicios de seguimiento y evaluación. Colombia*, Informe final de la Consultoría BID-RELAC.
- Red de Fondos Ambientales de Latinoamérica y el Caribe [<http://redlac.org/>], fecha de consulta: 29 de agosto de 2014.
- Redmexme [<http://redmexme.ning.com/>], consultada el 29 de agosto de 2014.
- Redlacme [<http://redlacme.org/>], fecha de consulta: 29 de agosto de 2014.

- Redcolme [<http://redcolme.ning.com/>], fecha de consulta: 1 de septiembre de 2014.
- Relac [<http://www.relacweb.org/>], fecha de consulta: 29 de agosto de 2014.
- Reyes, J.J. (2013). “Reto, conseguir la autonomía real del INEE”, *El Economista*, Sección Política, 2 de mayo [<http://eleconomista.com.mx/sociedad/2013/05/02/reto-conseguir-real-autonomia-inee>], fecha de consulta: 25 de agosto de 2014.
- Rodríguez, E. y E. Pasillas (2010), *Evaluación integral del desempeño de los programas federales de ayuda alimentaria y generación de capacidades 2010-2011* [http://www.coneval.gob.mx/Informes/Evaluacion/Integrales/Integrales%202010-2011/EI_10-11_Ayuda%20Alimentaria.pdf], fecha de consulta: 27 de octubre de 2014.
- Rodríguez, L. (2014), “Evaluación de la sustentabilidad en las áreas naturales protegidas en México”, tesis doctoral de medio ambiente y desarrollo de la Universidad Autónoma de Baja California Sur, México.
- Rojas, F. (2013), *Evaluación de desempeño organizacional del Consejo Nacional de Evaluación de la Política de Desarrollo Social* [http://www.coneval.gob.mx/Informes/Coordinacion/Evaluaciones_auditorias_externas/EDO_Resumen_Ejecutivo.pdf], fecha de consulta: 17 de noviembre de 2014.
- Roman, R. (2014). *Reglamentado Sistema Nacional de Evaluación y Gestión de Resultados –SINERGIA* [<http://redlacme.org/profiles/blogs/reglamentado-sistema-nacional-de-evaluacion-y-gesti-n-de>], fecha de consulta: 19 de julio de 2014.
- Sagarpa [<http://www.sagarpa.gob.mx/Paginas/default.aspx>], fecha de consulta: 19 de diciembre de 2014.
- Sagarpa (2011), Reglas de operación de los programas de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación, *Diario Oficial de la Federación*, 30 de diciembre de 2011.
- Sagarpa (2012a), *Evaluación de consistencia y resultados 2011-2012. Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural* [<http://www.sagarpa.gob.mx/programas2/evaluacionesExternas/Paginas/evalConsistencia-Resultados.aspx>], fecha de consulta: 5 de diciembre de 2014.
- Sagarpa (2012b), *Evaluación de consistencia y resultados 2011-2012. Programa de Apoyo al Ingreso Agropecuario: Procampo para Vivir Mejor* [<http://www.sagarpa.gob.mx/programas2/evaluacionesExternas/Paginas/evalConsistencia-Resultados.aspx>], fecha de consulta: 10 diciembre de 2014.
- Sagarpa (2013), *Programa sectorial de desarrollo agropecuario, pesquero y alimentario 2013-2018* [<http://www.sagarpa.gob.mx/asuntosinternacionales/cooperacioninternacional/Documents/Febrero%202014/Programa%20Sectorial%20de%20Desarrollo%20Agro%20Pesq%2020y%20Alim%20202013-2018.pdf>], fecha de consulta: 10 de diciembre de 2014.

- Sagarpa (2014), *Antecedentes* [<http://www.sagarpa.gob.mx/agricultura/Programas/proagro/procampo/Paginas/Antecedentes.aspx>], fecha de consulta: 1 de diciembre de 2014.
- Sánchez, M. (2013, 11 de mayo). “INEE constitucional, ¿autonomía para qué?”, *La Jornada*, Sección Opinión, 11 de mayo [<http://www.jornada.unam.mx/2013/05/11/opinion/019a2pol>], fecha de consulta: 25 de agosto de 2014.
- Sedema/CMM (2014), *Programa de Acción Climática de la Ciudad de México 2014-2020*, México: GDF.
- Sedesol (2008), *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007)* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 6 de noviembre de 2014.
- (2012a), *Metaevaluación 2007-2012* [http://www.2006-2012.sedesol.gob.mx/es/SEDESOL/Meta_Evaluacion_2007-2012], fecha de consulta: 15 de octubre de 2014.
- (2012b), *Aspectos susceptibles de mejora del Programa de Desarrollo Humano Oportunidades* [http://www.coneval.gob.mx/rw/resource/Mecanismos_2011/SEDESOL/PI_PDHO.pdf], fecha de consulta: 7 de octubre de 2014.
- Semarnat [<http://www.semarnat.gob.mx/>], fecha de consulta: 13 de enero de 2015.
- (2012a), Programa de Mejora de la Gestión. México.
- (2012b), *Informe de avances del Programa Especial de Cambio Climático 2009-2012* [<http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/pecc12.pdf>], fecha de consulta: 7 de octubre de 2014.
- (2013), *Avances en la atención de los ASM de septiembre de 2013* [<http://www.semarnat.gob.mx/programas/evaluaciones/avances-asm-sep2013>], fecha de consulta: 7 de octubre de 2014.
- (2014), “A finales de 2015 todas las ANP contarán con programas de manejo”, *Noticias MVS*, 1 de noviembre [<https://lineadirectaportal.com/publicacion.php?noticia=215041>], fecha de consulta: 30 de noviembre de 2014.
- SEP [<http://www.sep.gob.mx/es/sep1/>], fecha de consulta: 19 de diciembre de 2014.
- (2009), *Reglas de operación del Programa Escuelas de Tiempo Completo* [<http://basica.sep.gob.mx/tiempocompleto/pdf/ReglasOperacion2009.pdf>], fecha de consulta: 29 de octubre de 2014.
- (2010), *Programa Escuelas de Tiempo Completo. Documento de posicionamiento institucional* [http://www.sep.gob.mx/work/models/sep1/Resource/a69b6ff4-5c29-4565-ad3f-a8860965b055/doc_posic_petc_09_10.pdf], fecha de consulta: 29 de octubre de 2014.
- (2014a), *Acuerdo número 05/06/14 por el que se emiten los lineamientos de operación del Programa Escuelas de Excelencia para Abatir el Rezago Educativo* [http://dof.gob.mx/nota_detalle.php?codigo=5349156&fecha=18/06/2014], fecha de consulta: 21 de noviembre de 2014.

- (2014b), *Seguimiento a aspectos susceptibles de mejora clasificados como específicos. Programa Escuelas de Calidad* [http://www.sep.gob.mx/work/models/sep1/Resource/4660/1/images/11_310_S-029_DT.pdf], fecha de consulta: 7 de diciembre de 2014.
- SHCP (s/f), *Inventario de aspectos susceptibles de mejora con registro de información a septiembre de 2011* [www.hacienda.gob.mx/EGRESOS/sitio_pbr/evaluacion/ASM_bib/inventario_asm.pdf], fecha de consulta: 20 de octubre de 2014.
- (2012a), *Evaluación del sistema de presupuesto por resultados 2012* [http://www.siap.gob.mx/wp-content/pdf/programas_plan_desarrollo/ayuda.pdf], fecha de consulta: 14 de octubre de 2014.
- (2012b), *Catálogo de programas sujetos a Reglas de Operación. México, Distrito Federal* [http://www.shcp.gob.mx/EGRESOS/PEF/programacion/programacion2012/anexo_11_mppp_2012.pdf], fecha de consulta: 20 de noviembre de 2014.
- Silva, P. y ResilienceMonitor (2013), *Monitoreo y evaluación de medidas y proyectos para la adaptación al cambio climático en el sector agropecuario de la región Andina*. Informe de consultoría, ResilienceMonitor, España.
- SIMEC (2013), *Metadatos de indicadores estratégicos 2010-2013* [<https://simec.conanp.gob.mx/ind2013.php>], fecha de consulta: 12 de diciembre de 2014.
- Sinergia [<https://sinergia.dnp.gov.co/PortalDNP/default.aspx>], fecha de consulta: 13 de octubre de 2014.
- Skoufias, E. y J. Shapiro (2006), *Evaluating the impact of Mexico's Quality Schools Program: The pitfalls of using nonexperimental data*. Washington, DC: World Bank [<https://openknowledge.worldbank.org/bitstream/handle/10986/9010/wps4036.pdf?sequence=1>], fecha de consulta: 19 de diciembre de 2014.
- Société Française de l'Évaluation [<http://www.sfe-asso.fr/>], fecha de consulta: 19 diciembre de 2014.
- Solano, A. (s/f), “Propuesta sistema de evaluación programa nacional de promoción y protección de la salud”, tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- Soto, J.M. (2012), *Metaevaluación 2007-2012 del Programa de Desarrollo Humano Oportunidades. Informe final* [http://www.2006-2012.sedesol.gob.mx/work/models/SEDESOL/EvaluacionProgramasSociales/2012/META_EV_2012/PDHO_Metaevaluacion.pdf], fecha de consulta: 12 de diciembre de 2014.
- Spenthehauer, V. (1998), “L'évaluation des politiques publiques, avatar de la planification,” tesis doctoral de ciencia política de la Universidad Pierre Mendès-France de Grenoble, Francia.

- SS (2005), *Sistema de Protección Social en Salud. Estrategia de evaluación* [http://www.salud.gob.mx/unidades/evaluacion/seguropopular/estrategia_evaluac.pdf], fecha de consulta: 25 de noviembre de 2014.
- Stufflebeam, D.L. (1999), *Program evaluation metaevaluation checklist (based on the Program Evaluation Standards)* [<http://www.wmich.edu/evalctr/checklists/>], fecha de consulta: 4 de Octubre de 2010.
- Székely, M. (2013), *Evaluación externa de los resultados del Consejo Nacional de Evaluación de la Política de Desarrollo Social* [http://www.coneval.gob.mx/Informes/Coordinacion/Evaluaciones_auditorias_externas/Informe_EER.pdf], fecha de consulta: 17 de noviembre de 2014.
- Talbot, C. (2010), *Performance in government. The evolving system of performance and evaluation measurement, monitoring, and management in the United Kingdom*, Evaluation Capacity Development Working Paper Series núm. 24, Washington, DC: The World Bank [http://siteresources.worldbank.org/INTEVACAPDEV/Resources/ecd_24.pdf], fecha de consulta: 28 de junio de 2014.
- Taylor, E. y A. Yúnez (2013), *¿Procampo: hay fugas en la tubería?* [<http://cee.colmex.mx/documentos/documentos-de-trabajo/2013/dt20131.pdf>], fecha de consulta: 29 de diciembre de 2014.
- The Evaluation Center of the Western Michigan University [<http://www.wmich.edu/evaluation/>], fecha de consulta: 25 de septiembre de 2014.
- The Evaluator's Institute [<http://tei.gwu.edu/>], fecha de consulta: 18 de septiembre de 2014.
- Treasury Board of Canada Secretariat [www.tbs-sct.gc.ca], fecha de consulta: 8 de septiembre de 2014.
- UAM-Xochimilco (2009), *Sí Calidad. Evaluación técnico-financiera y de proceso del sistema integral de salud 2008, Síntesis ejecutiva* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 31 de octubre de 2014.
- (2009), *Sí Calidad. Evaluación técnico-financiera y de proceso del sistema integral de salud 2008, informe final* [http://portal.salud.gob.mx/codigos/columnas/evaluacion_programas/pdf/EXT09_SICALIDAD_IF.pdf], fecha de consulta: 31 de octubre de 2014.
- UICM y UNEP-WCMC (2012), *Informe de la situación del medio ambiente en México* [<http://es.scribd.com/doc/142035350/Informe-de-La-Situacion-Del-Medio-Ambiente-2012#scribd>], fecha de consulta: 14 de febrero de 2013.
- United Kingdom Government [<https://www.gov.uk/>], fecha de consulta: 18 de septiembre de 2014.
- United Kingdom Government, Cabinet Office [<https://www.gov.uk/government/organisations/cabinet-office/>], fecha de consulta: 14 de julio de 2014.

- United Kingdom Government, Better Regulation Executive [<https://www.gov.uk/government/groups/better-regulation-executive>], fecha de consulta: 22 de septiembre de 2014.
- United Kingdom Evaluation Society [<http://www.evaluation.org.uk/>], fecha de consulta: 16 de julio de 2014.
- University of Cambridge, Libraries Gateway [<http://www.lib.cam.ac.uk/libraries/>], fecha de consulta: 24 de septiembre de 2014.
- University of Oxford, Search Oxford Libraries Online [http://solo.bodleian.ox.ac.uk/primo_library/libweb/action/], fecha de consulta: 24 de septiembre de 2014.
- US Government Accountability Office (2013), *Summary of GAO's performance and accountability report* [<http://www.gao.gov/assets/660/659739.pdf>], fecha de consulta: 13 de julio de 2014.
- Vargas, O. (s/f), "Diseño de evaluación del componente de asistencia social, del área de solidaridad, ejecutado por el Instituto Mixto de Ayuda Social, IMAS", tesis de la Universidad de Costa Rica [sibdi.ucr.ac.cr/nuevo/html/TFG1997-2002.pdf], fecha de consulta: 14 de febrero de 2014 (ya no está disponible).
- Vernon, R. (2009), *Evaluación de procesos del Programa Fomento y Desarrollo de las Culturas Indígenas (Profodeci) 2009* [<http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>], fecha de consulta: 28 de octubre de 2014.
- Wingate, L.A. (2009), "The program evaluation Standards Applied for Metaevaluation Purposes: Investigating interrater reliability and implications for use", tesis doctoral de Western Michigan University, Kalamazoo, MI.

Siglas y acrónimos

AA2030:	Agenda del Agua 2030
AEA:	American Evaluation Association
AECID:	Agencia Española de Cooperación Internacional para el Desarrollo
AEE:	Agencia Española de Evaluación
AEMA:	Agencia Europea del Medio Ambiente
AG:	Acuerdos de Gestión
AMPE:	Asociación Mexicana de Profesionales de la Evaluación
ANMEB:	Acuerdo Nacional para la Modernización de la Educación Básica
ANP:	Áreas Naturales Protegidas
ASF:	Auditoría Superior de la Federación
ASM:	Aspectos Susceptibles de Mejora
BGI:	Balance de la Gestión Integral
BID:	Banco Interamericano de Desarrollo
BM:	Banco Mundial
Cader:	Centro de Apoyo al Desarrollo Rural
CAM:	Conferencia Anual del Municipio
CBO:	Congress Budget Office
CDI:	Comisión Nacional para el Desarrollo de los Pueblos Indígenas
CEC:	Comité d'Évaluation et de Contrôle des Politiques Publiques
CEDRSSA:	Centro de Estudios para el Desarrollo Rural Sustentable y la Soberanía Alimentaria
Cemabe:	Censo de Escuelas, Alumnos y Maestros de Educación Básica y Especial 2013
CEPAL:	Comisión Económica para América Latina
CES:	Canadian Evaluation Society
CICCDF:	Comisión Interinstitucional de Cambio Climático del Distrito Federal
CIDE:	Centro de Investigación y Docencia Económicas AC.
CIDS:	Comisión Intersecretarial de Desarrollo Social
CIESAS:	Centro de Investigaciones y Estudios Superiores en Antropología Social
CIMAP:	Comité Interministériel de Modernisation de l'Action Publique
CIME:	Comité Interministériel de l'Évaluation
CLAD:	Centro Latinoamericano de Administración para el Desarrollo
CLEAR:	Center for Learning on Evaluation and Result
CMM:	Centro Mario Molina
CNCH:	Cruzada Nacional Contra el Hambre

INSTITUCIONALIZACIÓN DE PROCESOS DE EVALUACIÓN

CNDS:	Comisión Nacional de Desarrollo Social
CNE:	Conseil National d'Évaluation des Politiques Publiques
CRS:	Congressional Research Service
Colmex:	El Colegio de México
Colpos:	Colegio de Postgraduados
Conabio:	Comisión Nacional para el Conocimiento y Uso de la Biodiversidad
Conaculta:	Consejo Nacional para la Cultura y las Artes
Conade:	Comisión Nacional de Cultura Física y Deporte
Conafe:	Consejo Nacional de Fomento Educativo
Conafor:	Comisión Nacional Forestal
Conagua:	Comisión Nacional del Agua
Conaliteg:	Comisión Nacional de Libros de Texto Gratuitos
Conanp:	Comisión Nacional de Áreas Naturales Protegidas
Conasupo:	Comisión Nacional de Subsistencias Populares
Conavi:	Comisión Nacional de Vivienda
Coneval:	Consejo Nacional de Evaluación de la Política de Desarrollo Social
CNPSS:	Comisión Nacional de Protección Social en Salud
Coplamar:	Coordinación del Plan General de Zonas Deprimidas y Grupos Marginados
CRIM:	Centro Regional de Investigaciones Multidisciplinarias
CSE:	Conseil Scientifique de l'Évaluation
DDR:	Distritos de Desarrollo Rural
DeGEval:	Gesellschaft für Evaluation e.V.
DEPPES:	Dirección de Evaluación de Programas Prioritarios y Especiales de Salud
DF:	Distrito Federal
DGDGIE:	Dirección General de Desarrollo de la Gestión e Innovación Educativa
DGE:	Dirección General de Evaluación
DGED:	Dirección General de Evaluación del Desempeño
DGGIMAR:	Dirección General de Gestión Integral de Materiales y Actividades Riesgosas
DGPAIRS:	Dirección General de Política Ambiental e Integración Regional y Sectorial
DGPE:	Dirección General de Planeación y Evaluación
Diconsa:	Distribuidora Conasupo
DIF:	Sistema Nacional para el Desarrollo Integral de la Familia
Dipres:	Dirección de Presupuestos
DNP:	Departamento Nacional de Planeación
DOAAM:	Dirección de Operación de la Atención a los Adultos Mayores
DOF:	Diario Oficial de la Federación
ECB:	Evaluation Capacity Building
ECR:	Evaluación de Consistencia y Resultados

SIGLAS Y ACRÓNIMOS

EDHUCA:	Estímulos para el Desarrollo Humano y las Capacidades de los Adultos
EED:	Evaluación Específica de Desempeño
EES:	European Evaluation Society
ELACCM:	Estrategia Local de Acción Climática de la Ciudad de México
ENAP:	École Nationale d'Administration Publique
ENCC:	Estrategia Nacional de Cambio Climático
Encel:	Encuesta de Evaluación de los Hogares Rurales
ENNVIIH:	Encuesta Nacional sobre Niveles de Vida de los Hogares
ENOE:	Encuesta Nacional de Ocupación y Empleo
Ensanut:	Encuesta Nacional de Salud y Nutrición
EPN:	Evaluación de Programas Nuevos
EUA:	Estados Unidos de América
EvalPartners:	The International Evaluation Partnership Initiative
Evalúa DF:	Consejo de Evaluación del Desarrollo Social del Distrito Federal
FANP:	Fondo para Áreas Naturales Protegidas
FAO:	Food and Agriculture Organization of the United Nations
Fidegar:	Fideicomiso para la Educación Garantizada
Flasco:	Facultad Latinoamericana de Ciencias Sociales
FMCN:	Fondo Mexicano para la Conservación de la Naturaleza
FNDE:	Fond National pour le Développement de l'Évaluation
Foceval:	Fortalecimiento de las Capacidades en Evaluación
FODA:	Fortalezas, Oportunidades, Debilidades y Amenazas
GAO:	General Accounting Office, Government Accountability Office
GDF:	Gobierno del Distrito Federal
GEI:	Gases de efecto invernadero
Gepanh:	Grupo de Investigación en Evaluación de Planes, Programas y Proyectos en el Campo de la Alimentación y Nutrición Humana
Gesoc:	Gestión Social y Cooperación AC.
GIZ:	Gesellschaft für Internationale Zusammenarbeit
GPRAMA:	Government Performance and Results Act Modernization Act
GRAEP:	Grupo de Responsables en Análisis y Evaluación de Programas
IAAM:	Instituto para la Atención de los Adultos Mayores
IDS:	Institute of Development Studies
ILE:	Interrupción Legal del Embarazo
IMAS:	Instituto Mixto de Ayuda Social
Inca Rural:	Instituto Nacional para el Desarrollo de Capacidades del Sector Rural
IMCO:	Instituto Mexicano para la Competitividad A.C.
IMSS:	Instituto Mexicano del Seguro Social

Indep:	Índice de Desempeño de los Programas Públicos Federales
INE:	Instituto Nacional de Ecología
INEA:	Instituto Nacional para la Educación de los Adultos
INECC:	Instituto Nacional de Ecología y Cambio Climático
INEE:	Instituto Nacional para la Evaluación de la Educación
INEGI:	Instituto Nacional de Estadística y Geografía
Infe:	Ley Federal de la Infraestructura Física Educativa
INGC:	Instituto Nacional de Gestión de Calamidades
INI:	Instituto Nacional Indigenista
Inifed:	Instituto Nacional de la Infraestructura Física Educativa
Inmujeres-DF:	Instituto Nacional de las Mujeres del Distrito Federal
INSP:	Instituto Nacional de Salud Pública de México
INVI:	Instituto de la Vivienda de la Universidad de Chile
IOCE:	International Organization for Cooperation in Evaluation
IPCC:	Panel Intergubernamental sobre Cambio Climático
IPE:	Institute for Program Evaluation
ISSSTE:	Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado
ITAM:	Instituto Tecnológico Autónomo de México
LAN:	Ley de Aguas Nacionales
LAP:	Loi sur l'Administration Publique
LDRS:	Ley de Desarrollo Rural Sustentable
LDSDF:	Ley de Desarrollo Social del Distrito Federal
LFPRH:	Ley Federal de Presupuesto y Responsabilidad Hacendaria
LGCC:	Ley General de Cambio Climático
LGDS:	Ley General de Desarrollo Social
LGE:	Ley General de Educación
LGEEPA:	Ley General de Equilibrio Ecológico y de Protección al Ambiente
LISMHDF:	Ley de Igualdad Sustantiva entre Mujeres y Hombres en el DF
LOAPF:	Ley Orgánica de la Administración Pública Federal
LOLF:	Loi Organique relative aux Lois de Finances
MDL:	Mecanismo de desarrollo limpio
MEC:	Mission de l'Evaluation et de Contrôle
MEP:	Ministerio de Educación Pública
MGEE:	Modelo de Gestión Educativa Estratégica
Mideplan:	Ministerio de Planificación Nacional y Política Económica; Ministerio de Planificación y Cooperación
MIR:	Matriz de Indicadores para Resultados
MIT:	Massachusetts Institute of Technology

SIGLAS Y ACRÓNIMOS

MML:	Matriz de Marco Lógico
MRV:	Medible, Reporte y Verificable
M&E:	Monitoreo y Evaluación
NAO:	National Audit Office
NOM:	Norma Oficial Mexicana
NPM:	New Public Management
OIT:	Organización Internacional del Trabajo
OMB:	Office of Management and Budget
OMS:	Organización Mundial de la Salud
OPS:	Organización Panamericana de la Salud
OSC:	Organizaciones de la Sociedad Civil
OTA:	Office of Technology Assessment
PAC:	Programa de Ampliación de Cobertura en Salud
PACCM:	Programa de Acción Climática de la Ciudad de México
PAE:	Programa Anual de Evaluaciones
PAL:	Programa de Apoyo Alimentario
PAPIR:	Programa de Apoyo a los Proyectos de Inversión Rural
PAR:	Programa de Abasto Rural
PARE:	Programa para Abatir el Rezago Educativo
PAREB:	Programa para Abatir el Rezago en Educación Básica
PART:	Program Assessment Rating Tool
PASL:	Programa de Abasto Social de Leche
PCyM:	Programa de Conservación y Manejo
PDCITYER:	Programa de Desarrollo de Capacidades, Innovación Tecnológica y Extensionismo Rural
PDHO:	Programa de Desarrollo Humano Oportunidades
PDSDF:	Programa de Desarrollo Social del Distrito Federal
PEC:	Programa Escuelas de Calidad
PEF:	Presupuesto de Egresos de la Federación
PEI:	Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras
PFDCI:	Programa Fomento y Desarrollo de Culturas Indígenas
PICC:	Panel Intergubernamental de Cambio Climático
Pider:	Programa de Inversiones Públicas para el Desarrollo Rural
PMyC:	Programas de Manejo y Conservación
PNANP:	Programa Nacional de Áreas Naturales Protegidas
PND:	Plan Nacional de Desarrollo
PNUD:	Programa de las Naciones Unidas para el Desarrollo
PPBS:	Planning, Programming and Budgeting System

PNUMA:	Programa de las Naciones Unidas para el Medio Ambiente
PRD:	Partido de la Revolución Democrática
Procampo:	Programa de Apoyo al Ingreso Agropecuario para Vivir Mejor
Procodes:	Programa de Conservación para el Desarrollo Sostenible
Prodesca:	Programa de Desarrollo de Capacidades en el Medio Rural
Profemor:	Programa de Fortalecimiento de Empresas y Organización Rural
Profepa:	Procuraduría Federal de Protección al Ambiente
Progesa:	Programa de Educación, Salud y Alimentación
Promanp:	Programa de Manejo de Áreas Naturales Protegidas
Promoeva:	Programa de Monitoreo y Evaluación del Desempeño Gubernamental 2010-2012
Promovi:	Programa de Monitoreo Biológico en Áreas Naturales Protegidas
Pronabes:	Programa Nacional de Becas para la Educación Superior
Pronasol:	Programa Nacional de Solidaridad
Provicom:	Programa de Vigilancia Comunitaria en Áreas Naturales Protegidas y Zonas de Influencia
PT:	Programa de Trabajo 2001-2006
RCB:	Rationalisation des Choix Budgetaires
Redcolme:	Red de Monitoreo y Evaluación de Política Pública
Redlacme:	Red Latinoamericana y del Caribe de Monitoreo y Evaluación
Redmexme:	Red Mexicana de Monitoreo y Evaluación
Relac:	Red de Seguimiento, Evaluación y Sistematización en América Latina y el Caribe
ReQUE:	Réseau québécois des universitaires en évaluation
RFE:	Réseau Francophone d'Évaluation
Rinace:	Red Iberoamericana de Investigación sobre Cambio y Eficacia Escolar
RLDSDF:	Reglamento de la Ley de Desarrollo Social del Distrito Federal
RO:	Reglas de Operación
RTP:	Red de Transporte de pasajeros
SACM:	Sistema de Aguas de la Ciudad de México
Sagar:	Secretaría de Agricultura, Ganadería y Desarrollo Rural
Sagarpa:	Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación
SAM:	Sistema Alimentario Mexicano
SCG:	Sistema de Control de Gestión
SDSDF:	Secretaría de Desarrollo Social
SECG:	Sistema de Evaluación y Control Gubernamental
Secodam:	Secretaría de la Contraloría y el Desarrollo Administrativo
Secogef:	Secretaría de la Contraloría General de la Federación
SED:	Sistema de Evaluación del Desempeño
Sedatu:	Secretaría de Desarrollo Agrario, Territorial y Urbano

SIGLAS Y ACRÓNIMOS

Sedema:	Secretaría del Medio Ambiente del GDF
Sedesol:	Secretaría de Desarrollo Social
Seduvi:	Secretaría de Desarrollo Urbano y Vivienda
Semarnap:	Secretaría de Medio Ambiente, Recursos Naturales y Pesca
Semarnat:	Secretaría de Medio Ambiente y Recursos Naturales
SEN:	Sistema Educativo Nacional
Senacatri:	Servicio Nacional de Capacitación y Asistencia Técnica Rural Integral
Senara:	Servicio Nacional de Aguas Subterráneas, Riego y Avenamiento
SEP:	Secretaría de Educación Pública
Sepsa:	Secretaría de Planificación del Sector Agropecuario
Sesa:	Servicios Estatales de Salud
Setravi:	Secretaría de Transportes y Vialidad
SFP:	Secretaría de la Función Pública
SHCP:	Secretaría de Hacienda y Crédito Público
Sicalidad:	Sistema Integral de Calidad en Salud
Simactic:	Sistema de Monitoreo y Acompañamiento de las Escuelas de Tiempo Completo
SIMEC:	Sistema de Información, Monitoreo y Evaluación para la Conservación
Simep:	Sistema de Seguimiento de Metas Presidenciales
Sinacatri:	Sistema Nacional de Capacitación y Asistencia Técnica Rural Integral
Sinap:	Sistema Nacional de Áreas Protegidas
Sine:	Sistema Nacional de Evaluación
Sinergia:	Sistema Nacional de Evaluación de Resultados de la Gestión Pública
Sipromoeva:	Sistema del Programa de Monitoreo y Evaluación del Desempeño Gubernamental
Sisdeval:	Sistema Nacional de Evaluaciones
SiSEP:	Sistema de Evaluación y Seguimiento de Proyectos
Sismeg:	Sistema de Seguimiento a Metas de Gobierno
SFE:	Société Française d'Évaluation
SNDIF:	Sistema Nacional para el Desarrollo Integral de la Familia
SNEE:	Sistema Nacional de Evaluación Educativa
SNTE:	Sindicato Nacional de Trabajadores de la Educación
SPP:	Secretaría de Programación y Presupuesto
SPPA:	Subsecretaría de Planeación y Política Ambiental
SPSS:	Sistema de Protección Social en Salud
SOEP:	Société Québécoise d'Évaluation de Programme
SS:	Secretaría de Salud
SSDF:	Secretaría de Salud del Distrito Federal
STE:	Sistema de Transporte Eléctrico
SURI:	Sistema Único de Registro de Información

INSTITUCIONALIZACIÓN DE PROCESOS DE EVALUACIÓN

TLCAN:	Tratado de Libre Comercio de América del Norte
TNC:	The Nature Conservancy
TPCP:	Transporte público colectivo de pasajeros
TR:	Términos de referencia
UA-FAO:	Unidad de Apoyo FAO
UAM-Xochimilco:	Universidad Autónoma Metropolitana, Unidad Xochimilco
UE:	Unión Europea
UICN:	Unión Internacional para la Conservación de la Naturaleza
UKES:	United Kingdom Evaluation Society
UNAM:	Universidad Nacional Autónoma de México
UNEG:	United Nations Evaluation Group
UNEP:	United Nations Environment Programme
Unesco:	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
Unicef:	Fondo de las Naciones Unidas para la Infancia
VIH:	Virus de inmunodeficiencia humana
WCMC:	World Conservation Monitoring Centre
WWF:	Fondo Mundial para la Conservación de la Naturaleza

Índice de cuadros y gráficas

PRIMERA PARTE

Institucionalización y profesionalización en evaluación

Cuadro 1: Estados Unidos: principales programas académicos por universidades	37
Cuadro 2: Reino Unido: principales programas académicos por universidad	48
Cuadro 3. Francia: principales eventos realizados anualmente por la SFE 2011-2014	56
Cuadro 4. Francia: oferta universitaria de formación en evaluación	65
Cuadro 5. Oferta universitaria de formación en evaluación en Canadá, exceptuando la provincia de Quebec	76
Cuadro 6. Oferta universitaria de formación en evaluación en Quebec, Canadá	78
Cuadro 7. Colombia: evaluaciones contratadas por el DNP	84
Cuadro 8. Chile: Sistema de Evaluación y Control de Gestión	102
Cuadro 9. Chile: diseño operativo de las evaluaciones	103
Cuadro 10. Chile: principales programas de formación en evaluación	105
Cuadro 11. México: evaluaciones programadas y realizadas por el Coneval 2007-2011	112
Cuadro 12. Evaluaciones realizadas por Evalúa DF 2008-2013	120
Cuadro 13. Evaluaciones de programas de desarrollo social 2007-2014	128
Cuadro 14. Evaluaciones de programas de educación 2007-2014	140
Cuadro 15. Evaluaciones de programas de medio ambiente por tipo 2007-2012	147
Cuadro 16. Evaluaciones de programas de medio ambiente 2013-2014	147
Cuadro 17. Evaluaciones de programas de salud 2007-2014	162
Cuadro 18. Evaluaciones de programas de Sagarpa 2001-2013	171
Cuadro 19. Síntesis de característica de la evaluación por países	175
Cuadro 20. Síntesis de las características de la evaluación por sectores mexicanos	178

SEGUNDA PARTE

Metaevaluación: calidad de las evaluaciones aplicadas en México 2007-2014

Cuadro 21. Características por tipo de evaluación	191
Cuadro 22. Índice de desempeño de algunos programas públicos federales 2014	195
Cuadro 23. “Metaevaluaciones” 2007-2012 coordinadas por la Sedesol	197
Cuadro 24. Selección de evaluaciones a metaevaluar a nivel federal	200
Cuadro 25. Selección de evaluaciones a metaevaluar del GDF	201
Cuadro 26. Análisis de líneas e indicadores propuestos por la Evaluación de medio término del PNANP	263
Cuadro 27. Distribución del tiempo de trabajo por evaluación y etapas	323
Cuadro 28. Instrumentos de recopilación de información utilizados	326
Cuadro 29. Recomendaciones emitidas por Evalúa DF en 2011-2012	334
Cuadro 30. Respuesta de las dependencias a las recomendaciones	336
Cuadro 31. Principales resultados de las metaevaluaciones federales	352
Gráfica 1. Costos de las evaluaciones federales	353
Cuadro 32. Costo y calidad de las evaluaciones federales	355

TERCERA PARTE

Utilización de las evaluaciones

Cuadro 33. Principales factores explicativos de la utilización de las evaluaciones (autores y año de publicación)	380
Cuadro 34. Número de recomendaciones y de ASM en la generación de evaluaciones 2012-2013	387
Cuadro 35. Número de programas evaluados, de instancias responsables y de ASM 2008-2013	388
Cuadro 36. Distribución de ASM por tipos 2008-2013	389
Cuadro 37. Datos complementarios sobre las evaluaciones realizadas y los ASM emitidos	390
Cuadro 38. Seguimiento de los ASM identificados desde 2010 a la fecha	392
Cuadro 39. ASM cancelados 2013-2014	394
Cuadro 40. ASM dados de baja 2013-2014	394

ÍNDICE DE CUADROS Y GRÁFICAS

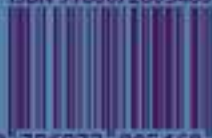
Cuadro 41. Reconocimiento de buenas prácticas en el uso de los resultados de monitoreo y evaluación	398
Cuadro 42. Mejoras instrumentadas por evaluación y programa	400
Cuadro 43. Presupuesto por programas de la Sedesol 2007-2012	405
Cuadro 44. Número de ASM por tipo en la Sedesol	409
Cuadro 45. Definición y seguimiento de ASM de Oportunidades 2007-2010	410
Cuadro 46. Número de ASM por tipo en la SEP	416
Cuadro 47. Número de ASM por tipo en la Semarnat	419
Cuadro 48. Número de ASM por tipo en la SS	422
Cuadro 49. Número de ASM por tipo en la Sagarpa	424
Cuadro 50. Recomendaciones 2009-2011	427
Cuadro 51. Estado de las recomendaciones 2009 al 30 de junio de 2011	427
Cuadro 52. Gasto social ejercido en los principales organismos públicos del GDF 2008-2012	433
Cuadro 53. Enfoques de la evaluación	435
Cuadro 54. Prácticas de evaluación	435

Institucionalización de procesos de evaluación. Calidad y utilización de sus resultados, coordinado por Myriam Cardozo Brum, número 54 de la Colección Teoría y análisis de la DCSH de la UAM-Xochimilco, terminó de imprimirse el 29 de noviembre de 2015. Edición: Vákşu editores. Impresión: mc editores, Selva 53-204, col. Insurgentes Cuicuilco, 04530 México, D.F., tel. 5665 7163 [mceditores@hotmail.com]. El tiro consta de mil ejemplares más sobrantes para reposición.

Con motivo del Año Internacional de la Evaluación (2015), promovido por The International Evaluation Partnership Initiative (EvalPartners) y asumido por organismos internacionales, gobiernos nacionales, fundaciones privadas y sociedades de evaluación² con la finalidad de realizar múltiples actividades tendientes a poner a la evaluación en el centro del debate público, y considerando la importancia adquirida por los procesos de evaluación de las políticas y programas gubernamentales, la Universidad Autónoma Metropolitana, Unidad Xochimilco, y el Centro de Estudios Sociales y de Opinión Pública de la H. Cámara de Diputados, han querido contribuir con este libro dedicado a actualizar tanto las discusiones teóricas como las experiencias prácticas, especialmente en el contexto mexicano.

La evaluación permite conocer el nivel de desempeño de políticas y programas (resultados e impactos), explicarlo tomando en consideración las características de su diseño e instrumentación pero también del contexto político, social y económico concreto en que se inserta, y valorarlo a la luz de un conjunto de criterios como la pertinencia, la congruencia, la eficacia, la eficiencia, la calidad, la equidad, entre otros. Para ello requiere de: a) un proceso de institucionalización que facilite el desarrollo de una investigación aplicada e interdisciplinaria por parte de un equipo con formación y experiencia tanto en sus métodos como en torno al problema que es objeto de la acción pública a analizar; b) mecanismos que garanticen la calidad de los trabajos y la factibilidad de sus recomendaciones; y c) la utilización de sus resultados para satisfacer las necesidades antes planteadas. Estos tres aspectos constituyen los temas tratados en cada una de las partes del presente libro.

ISBN 9786072805460



9 786072 805460



Publicaciones