

La evaluación de programas en Chile: análisis de una muestra de programas públicos evaluados*

Mauricio Olavarría Gambi

Introducción

¿Qué efectos han tenido las evaluaciones de programas públicos emprendidas en Chile sobre las decisiones de los poderes públicos y la gestión de los programas? La evaluación de programas ha ido siendo vista con creciente interés por los máximos líderes del país como una herramienta que permite mejorar tanto la propia gestión de los programas como las decisiones de política sobre los programas públicos. Así, los gobiernos de la ex presidenta Bachelet y del presidente Piñera han emprendido serios esfuerzos por instalar en Chile una “Agencia de Evaluación de Políticas Públicas”. En este contexto, este estudio busca aportar evidencia de primera fuente acerca del funcionamiento cierto de la evaluación de programas en el país.

Aunque la práctica de la evaluación de las intervenciones del Estado tiene un antecedente remoto en la preocupación por el adecuado uso de los recursos públicos, la evaluación de programas, como actividad sistemática de aplicación de metodologías de investigación de ciencias sociales, emprendida por una dependencia del gobierno central, data de la segunda mitad de la década de 1990 (Olavarría y Figueroa, 2010). En 1996, luego de un acuerdo entre el Poder Ejecutivo y el Congreso, orientado a generar mayor información de calidad para la discusión acerca de los proyectos anuales de ley de presupuesto, se estableció el Programa de Evaluación de Programas en la Dirección de Presupuestos (DIPRES) del Ministerio de Hacienda. El Programa se inició con evaluaciones basadas en el “Juicio de Expertos”, con paneles integrados por funcionarios públicos y profesionales externos, y a partir de 1998 se han efectuado solo con profesionales externos, que concursan públicamente y se inscriben en un registro de consultores del Estado. En el 2001 se incorporó la modalidad de Evaluación de Impacto, que es desarrollada por consultores externos seleccionados mediante licitación pública.

El modo en que surgió la práctica de evaluación de las intervenciones públicas en Chile llevó a orientarla a la verificación del cumplimiento de metas de control del gasto público y a organizarla alrededor del ciclo presupuestario. La finalidad de ello es que los resultados de las evaluaciones pudieran tener un correlato en las decisiones políticas sobre asignación de recursos y que los evaluados tuvieran suficientes incentivos para hacer un “adecuado uso de los recursos públicos” asignados.

Dado el modo como se ha establecido y ejercido la evaluación de programas en Chile, lo que interesa saber es si esta práctica se ha constituido en instrumento útil para el proceso de toma de decisiones del Estado chileno y para el diseño, formulación e implementación de políticas públicas, y si de ello surge evidencia que aconseje una distinta organización de la práctica de la evaluación. De este modo, entonces, el estudio recoge evidencia acerca de la utilidad de la práctica chilena de evaluación de programas, y si una eventual Agencia de Evaluación de Políticas Públicas debería recoger la práctica y organización que ha imperado entre 1996 y 2012, o si, alternativamente, debería organizarse y estructurar el trabajo de manera distinta.

El documento se organiza en seis secciones. Luego de esta introducción, se exponen los conceptos básicos acerca de la relación entre política pública y programa público, la noción de evaluación de programas y los métodos principales usados en la práctica chilena. La sección siguiente explica los aspectos metodológicos del trabajo. La cuarta sección presenta la evidencia recogida,

Recibido: 12-06-2012. Aceptado: 08-09-2012.

(*) En este trabajo han colaborado los siguientes profesionales: Cesia Arredondo, Carlos Andrade y Javier Pino. A ellos les expreso mi reconocimiento y agradecimiento por su aporte. El presente estudio es parte del proyecto “Apoyo al diseño e instalación de la Agencia de Evaluación de Programas e Inversiones Públicas” impulsado y financiado por el Banco Interamericano de Desarrollo.

seguida de una sección de discusión e interpretación de la evidencia. La sección final exhibe las conclusiones e implicancias que fluyen del análisis.

Conceptos básicos

La unidad de análisis del estudio es la evaluación efectuada a programas públicos -a través de la Evaluación de Programas Gubernamentales (EPG) y Evaluación de Impacto (EI)- que ponen en funcionamiento intervenciones de política pública.

Un programa público es una unidad de trabajo que operacionaliza una política pública y que, a su vez, se desagrega en proyectos específicos que materializan la intervención en los distintos ámbitos -o componentes- abordados por el programa. Es decir, la acción pública reconoce una jerarquía de sus intervenciones en la que la política pública es el nivel superior, que identifica y explicita el objetivo político buscado; el programa es el nivel *meso*, que operacionaliza el objetivo político para una determinada área de actuación; y el proyecto representa el nivel micro, que materializa acciones específicas orientadas a cumplir metas asignadas al programa.

Así, mientras una política pública “puede ser definida como una intervención del Estado, expresada en una decisión o conjunto de decisiones de una autoridad pública, que considera un análisis técnico-racional para un tema determinado y una finalidad específica, que sigue un procedimiento formal, todo lo cual se da en el contexto de un intenso proceso político de confrontación y articulación de intereses” (Olavarría, 2010: 12); un programa sería “un conjunto organizado de actividades, proyectos, procesos o servicios orientados a la consecución de objetivos específicos” (ONU, 1984: 15-16), o un conjunto de recursos y actividades dirigidas hacia el logro de uno o más objetivos comunes, usualmente bajo la dirección de un administrador o un equipo de gestión (Wholey ...[et al], 2004). De este modo, entonces, el programa establece las prioridades de intervención, identifica y ordena los proyectos, define el marco institucional y asigna los recursos que se utilizarán (Cohen y Franco, 2000).

En este contexto, la evaluación es entendida como la aplicación sistemática de métodos de investigación de ciencias sociales para juzgar y sopesar la conceptualización, diseño, implementación y utilidad de programas de intervención social (Rossi y Freeman, 1993). Para Subirats (1994: 139), en tanto, la evaluación es un “método de investigación sistemática sobre la configuración de un programa o política determinada, y sobre su implementación y efectividad”. De este modo, entonces, “la evaluación es una práctica que se da a lo largo del proceso de la política pública y que tiene por finalidad aportar información válida a la toma de decisiones acerca de los rendimientos de la intervención. (...) son investigaciones llevadas a cabo para afectar el desarrollo de las políticas públicas, para influir y dar forma al diseño e implementación de las intervenciones del Estado, para mejorar la administración de ellas, para identificar los resultados alcanzados, para estimar su contribución a las metas gubernamentales y la eficiencia, efectividad e impacto de los emprendimientos estatales en los que se expresan tales intervenciones” (Olavarría y Figueroa, 2010: 13).

Como se ha mencionado previamente, dos tipos de prácticas evaluativas han sido recurrentes en Chile: la EPG y la EI. Ambas se constituyen en el foco de este trabajo.

La EPG es una práctica de evaluación que se ha constituido con base en el uso de “Juicio de Expertos”. En ella, la DIPRES constituye un panel normalmente integrado por tres expertos¹ en el tema o problema abordado por el Programa Público a evaluar. El trabajo se desarrolla entre los meses de enero y julio de cada año, siguiendo una pauta elaborada por la DIPRES, y según la cual los consultores deben entregar cinco informes en diferentes momentos². El trabajo evaluativo de los consultores se centra en exponer la justificación del programa, el grado de cumplimiento de su fin y propósito, la producción de bienes y servicios, la cobertura y focalización, la satisfacción de los usuarios, las fuentes de financiamiento y el gasto, la eficiencia del gasto, y recomendaciones (DIPRES, 2012).

El Marco Lógico del programa es el instrumento que guía la evaluación, pues, con él a la vista, el

trabajo consiste en comparar la implementación con lo expresado en el diseño. La información recogida por los expertos es aportada por el propio programa evaluado, el organismo público que alberga al programa o la DIPRES, aunque también existe la posibilidad de efectuar estudios complementarios o recabar información adicional.

Las evaluaciones de “Juicio de Expertos” pueden presentar sesgos que, en algunos casos, llegan a ser considerables. Es normal que entre los expertos haya desacuerdos sobre los temas de su dominio, por diversidad de enfoques conceptuales, aproximaciones metodológicas, conclusiones de sus propios estudios anteriores, entre otros aspectos; todo lo cual influye en el modo en que perciben el programa, en el juicio que se forman acerca de su funcionamiento y en el informe de evaluación que presentan (Rossi y Freeman, 1993). Aunque es necesario tener presente sus limitaciones, la evaluación de expertos es recurrentemente una opción recomendable por la premura con que se requiere disponer de una “opinión” acerca del programa, la baja disponibilidad de presupuesto para la evaluación u otras restricciones.

El tipo de evaluación que se efectúa a través de la EPG es aquel que la literatura especializada identifica como Monitoreo o Seguimiento³. Esta es una “evaluación intermedia” que ocurre durante la ejecución del programa o proyecto y cuyo propósito es poder identificar si hay circunstancias que estén impidiendo que la intervención marche como se preveía o cumpla las metas que se habían presupuestado para los distintos momentos de la ejecución. Este tipo de evaluación supone la existencia de un plan, que establezca metas e indicadores con el cual comparar la ejecución. Su propósito es identificar las causas de los problemas que están afectando a la implementación del programa, e informar a los responsables de la gestión para que estos implementen las acciones correctivas, que pueden apuntar a correcciones del diseño o correcciones en la gestión del programa. Así, el monitoreo puede ser visto como una actividad orientada a ajustar la implementación del programa con su diseño. Sin un adecuado monitoreo o seguimiento es imposible saber si la intervención se está encaminando a los resultados que se esperan.

Valadez y Bamberger (1994) definen el monitoreo como una actividad gerencial interna y continua cuyo propósito es asegurar que el programa alcance sus objetivos, en el tiempo previamente establecido y conforme al presupuesto asignado. De este modo, el monitoreo involucra la provisión de información de retroalimentación sobre los avances y problemas detectados durante la implementación. Las actividades en las que se centra el monitoreo, para estos autores, son el análisis de la obtención y asignación de recursos, producción y entrega del servicio, y el registro de los costos. Cohen y Franco (2000) señalan que el monitoreo o seguimiento es el examen continuo o periódico que efectúa la administración, en todos sus niveles jerárquicos, acerca de la manera en que se está ejecutando una actividad. Para Rossi y Freeman (1993) el monitoreo es el examen sistemático de la cobertura del programa y su implementación, distribución y entrega. La evaluación de la cobertura consiste en estimar en qué medida el programa llega a su población objetivo. El análisis de la implementación consiste en la determinación de la congruencia entre lo establecido en el plan respecto del otorgamiento de los servicios del programa y los que realmente se están entregando.

El otro tipo de práctica evaluativa recurrentemente aplicada en Chile es la Evaluación de Impacto (EI), que se ha ejecutado a través de diseños cuasi-experimentales. La pregunta que una evaluación de impacto intenta responder es ¿qué cambios ocurridos en la realidad pueden ser razonablemente atribuidos al programa? El objetivo básico de este tipo de evaluación es producir una estimación de los efectos de una intervención, que no estén contaminados por la influencia de otros procesos y eventos que también pueden haber afectado la conducta o condiciones de implementación de la intervención que está siendo evaluada.

Dado que, con frecuencia, la decisión de evaluar el impacto de un programa se toma cuando este ya se está implementando, la opción metodológica normal es el diseño cuasi-experimental, que es un diseño robusto y muy frecuentemente aplicado. El punto de cuidado es si el grupo de comparación que

se ha elegido es suficientemente comparable al grupo que ha recibido el tratamiento en todas las características y experiencias relevantes para el resultado del programa o proyecto que está siendo evaluado⁴.

En el contexto del presente trabajo, y dada la conceptualización expuesta, la cuestión pasa a ser, entonces, el efecto que ha tenido sobre las decisiones de los poderes públicos y gerentes de programas la aplicación de estos conceptos en la práctica chilena.

Métodos y datos

Como se ha mencionado, la unidad de análisis del estudio se constituye con programas públicos que han sido sometidos a la EPG, mediante el método de “Juicio de Expertos” o a la EI. El universo de programas evaluados entre los años 1999 y 2010 es de 255 programas en el contexto del Programa de Evaluación de Programas Gubernamentales (EPG), 85 evaluaciones de impacto (EI) y nueve evaluaciones de programas nuevos (EPN)⁵ (Arenas y Berner, 2010: 47).

La Evaluación de Programas Nuevos (EPN) no ha sido incluida en el universo de programas desde los cuales se dedujo la muestra, porque al momento de realización de este trabajo no había información disponible sobre la evaluación de ellos.

De este universo se ha seleccionado una muestra con la finalidad de comprender cómo se da el proceso de evaluación de programas públicos en la práctica y su efecto en las decisiones de los poderes públicos y la gestión de los programas. Para la identificación de los casos de análisis se aplicaron criterios de muestreo intencionado y de saturación teórica. Por el primero se busca recoger información de fuentes directamente vinculadas al caso en estudio y cuyos testimonios permitan identificar categorías analíticas relevantes, en tanto que a través del segundo se busca evitar recoger información que ya se dispone o que no agregue valor analítico (Valles, 2007).

La muestra incluye 23 programas públicos evaluados (ver Cuadro 1). De ellos, siete han tenido más de una evaluación (hasta tres), cinco registran dos evaluaciones de EPG y tres han sido evaluados tanto a través de “Juicio de Expertos” como de EI.

Cuadro 1
Muestra de Programas Evaluados Analizados

Programa	Evaluaciones de Programas Gubernamentales	Año de la publicación	Evaluaciones de Impacto	Año de la publicación
Programa de atención a menores: Administración Directa	2	2000-2007	0	---
Programa Fortalecimiento de la Formación Inicial Docente	1	2001	0	---
Programa Orígenes	2	2004-2010	0	---
Programa de Prevención y Control del VIH/SIDA y Enfermedades de Transmisión Sexual (ETS)	2	2000-2010	0	---
Fondo Social	1	2001	0	---
Chile Barrio	1	2002	1	2007
Seguridad y Participación	1	2010	0	---
Plan Cuadrante	1	2007	0	---
Programa de Fomento CORFO	0	---	1	2009
Becas Presidente de la República (Enseñanza Media)	1	1999	1	2006
Programa Administración de Concesiones	1	2007	0	---
Programa de Mejoramiento de la Calidad y Equidad de la Educación Superior (MECESUP)	1	2004	0	---
Programa de Mejoramiento Urbano y Equipamiento Comunal (PMU)	2	1999-2011	1	2002
Programa para la Recuperación de Suelos Degradados	2	2000-2009	0	---
Programa Evaluación de Programas	0	---	1	2005
Programa Red Tecnológica Educativa Enlaces	0	---	1	2004
Fondo de Tierras y Aguas Indígenas	0	---	1	2002
Programa Nacional Mujeres Jefas de Hogar	0	---	1	2010
Programa Mejoramiento de la Gestión	0	---	1	2008
Programas de Empleo con Apoyo Fiscal	0	---	1	2004
Bonificación a la Contratación de Mano de Obra	0	---	1	2009
Programa de Educación y Capacitación Permanente Chilecalifica	0	---	1	2009
Chile Solidario	0	---	1	2006

Fuente: elaboración propia.

Los criterios utilizados para seleccionar la muestra son los siguientes:

a) Tamaño del programa evaluado: ha sido operacionalizado a través del gasto que ha significado para el Estado por el período evaluado, según lo reportado por el propio informe de evaluación del programa. Dado que la inversión está normalmente asociada a la prioridad en la agenda gubernamental, se han seleccionado evaluaciones de programas cuyo costo reportado es superior al promedio del universo.

b) Sector de actividad gubernamental: la muestra debe expresar un razonable equilibrio entre los distintos sectores de actividad del gobierno.

c) Importancia para la Agenda Pública: expresa la correspondencia entre las preocupaciones ciudadanas y las decisiones del Estado de evaluar su actividad. Así, según este criterio, las evaluaciones analizadas deben también incorporar, por un lado, a aquellos programas que abordan problemas públicos que han tenido alta prioridad en las percepciones ciudadanas, representadas a través de las encuestas de opinión pública del Centro de Estudios Públicos (CEP). Por otro, hace referencia a aquellos programas que han tenido más de una evaluación, pues ello muestra una preocupación del Estado acerca de si los resultados e impactos que exhibe el programa se corresponden con las expectativas que lo hicieron surgir y, en último término, con la satisfacción de aspiraciones ciudadanas.

d) Duda: expresa una percepción de anomalías o falta de efectividad en el programa evaluado. Se operacionaliza a través de informes públicos, investigaciones de la Contraloría General de la República y noticias o crónicas aparecidas en medios de comunicación que reportan alguna de las

situaciones señaladas.

La pregunta que da origen al estudio ha sido operacionalizada en las siguientes cuatro preguntas específicas:

- 1) ¿Utilizan los poderes públicos los resultados de las evaluaciones de programas para sus deliberaciones y decisiones?
- 2) ¿Qué efectos ha generado al interior de los organismos y en la gestión de programas los resultados de la evaluación que se les ha efectuado?
- 3) ¿Qué efectos han tenido las evaluaciones de programas emprendidas en Chile en las decisiones de autoridades superiores del Poder Ejecutivo?
- 4) ¿Cuál es la calidad de la evaluación de programas emprendida en Chile?

La pregunta N° 1 indaga sobre si la evaluación efectuada a programas ha sido utilizada en las discusiones que se dan para la conformación y aprobación de la ley de presupuesto, o en las deliberaciones que se dan entre los poderes públicos acerca de la creación de un programa, servicio público o de la marcha de la gestión del Estado. Es decir, esta pregunta busca aportar información acerca de la incidencia de la evaluación de programas en el debate que se da en el sistema político sobre las políticas públicas.

La pregunta N° 2 se orienta a recabar información acerca de si los encargados de la gestión de los programas públicos han recogido los resultados de la evaluación para mejorar diseños y prácticas, para establecer compromisos de rendimiento, reorientación del programa, o si más bien la evaluación ha sido vista meramente como el cumplimiento de un requerimiento formal del programa.

La pregunta N° 3 está dirigida a establecer si las autoridades superiores de los organismos encargados de la gestión de los programas utilizan los resultados de las evaluaciones para sus procesos de decisiones sobre el programa o sobre los organismos en referencia. La pregunta también apunta a identificar el efecto que las evaluaciones tienen sobre las decisiones de continuidad o de asignaciones presupuestarias a los programas o al organismo a cargo de su gestión.

La pregunta N° 4 pretende aportar información respecto de si las evaluaciones emprendidas incorporan metodologías reconocidas en la literatura especializada, si las técnicas analíticas empleadas llevan a reportar resultados insesgados, si el informe de evaluación expresa una rigurosa aplicación de los métodos de análisis y si la evaluación logra recoger los aspectos sustantivos y complejidades del programa.

La información a la que se acude para responder las cuatro preguntas identificadas es variada. Los esfuerzos para responder la pregunta N° 1 se centran en el análisis de la historia de las leyes de presupuesto de los años 2006 a 2011 (las únicas disponibles en el sitio web de la Biblioteca del Congreso Nacional). La finalidad es establecer si en el debate en sala o en el trabajo de comisiones de presupuesto las evaluaciones de programas fueron usadas, citadas, referidas o aludidas en la discusión de la partida presupuestaria que asignaba recursos al programa o al órgano encargado de su gestión. También se indaga si en la creación, modificación o cese de programas u órganos relacionados con el programa evaluado fueron consultadas esas evaluaciones.

Para responder las preguntas Nos. 2 y 3 se acude a una información dual. Por un lado, a los informes de “Balance de Gestión Integral” de los servicios que albergan los programas evaluados, que se hayan efectuado con posterioridad a la evaluación del programa analizada en este trabajo. Adicionalmente, se hicieron 13 entrevistas (ver Cuadro 2), en el mes de febrero de 2012, a responsables de los programas y autoridades de los servicios que albergan los programas al momento de la evaluación, además de parlamentarios, con la finalidad de recabar información acerca de las consecuencias que habrían generado los informes de evaluaciones sobre la gestión de los programas, su continuidad y asignaciones presupuestarias. La personas entrevistadas han sido seleccionadas en razón del rol desempeñado en algún(os) programa(s) incluido(s) en la muestra. Es de notar que, en varios casos, las personas entrevistadas también tuvieron relación con otros programas evaluados no incluidos

en la muestra de este estudio y en distintos roles (administrador de programa, autoridad gubernamental o parlamentario). La entrevista abordó -en los casos que correspondía- la visión de los entrevistados acerca de la práctica de evaluación de esos otros programas, cuestión que permitió recoger información amplia y de alto valor analítico. Las entrevistas fueron analizadas mediante la técnica del “análisis de contenido”.

Cuadro 2
Lista de Entrevistados

Entrevistado	Cargo al momento de la evaluación	Programa	Año evaluación
María Eugenia Martínez	Encargada PMU SUBDERE	Programa de Mejoramiento Urbano (PMU)	2011
Karina Doña	Directora Programa Orígenes	Orígenes	2010
Gonzalo Duarte	Diputado; ex Director Chile Barrio	Chile Barrio y otros	2007
Pilar Armanet	Jefa División Educación Superior MINEDUC	MECESUP y otros	2004
Laura Albornoz	Ministra SERNAM	Programa Nacional de Mujeres Jefas de Hogar y otros	2009
Virginia Astorga	Jefa de la Unidad de Formación Técnica, Programa <i>Chilecalifica</i>	Programa de Educación y Capacitación Permanente, Chilecalifica	2009
Eduardo Abedrapo	Subsecretario MIDEPLAN	Becas Presidente de la República / Chile Solidario y otros	2006
Andrés Palma	Ex Diputado; ex Director Chile Barrio	Chile Barrio y otros	2007
Alberto Cienfuegos	General Director Carabineros	Plan Cuadrante	2007
Carlos Álvarez	Vicepresidente Ejecutivo CORFO	Fomento: FOCAL, PROFO, PDP, PTI, PI, PEL, FAT y otros	2007
Tamara Jieri	Consultora SUR Profesionales	Orígenes y otros	2010
Laura Echeverría	Encargada Técnica Programa	Programa Nacional de Mujeres Jefas de Hogar	2009
Eugenio San Martín	Director SENAME	Programa de Atención a Menores: Administración Directa	2007

Fuente: elaboración propia.

Para analizar la calidad de la información de las evaluaciones de impacto, el estudio ha construido una escala del rigor metodológico de la evaluación (Cuadro 3). Este instrumento se inspira en la conocida “Escala de Maryland”, construida por Sherman ...[et al] (1998) para analizar la calidad de las evaluaciones efectuadas sobre intervenciones de prevención delictual e informar al Congreso estadounidense acerca de lo que funciona, lo que no funciona y lo que es prometedor en la prevención del delito en Estados Unidos. Basado en ese instrumento, la “Escala de Rigor Metodológico en las Evaluaciones de Impacto de los Programas” efectuadas en Chile considera las siguientes categorías, donde 1 es la de menor rigor y 5 es la de mayor rigor.

Cuadro 3
Escala de Rigor Metodológico de las Evaluaciones de Programas en Chile

Categoría	Significado
1	Análisis de archivos, entrevistas a funcionarios y usuarios del programa, conclusiones generales e impresiones acerca de la efectividad del programa, o análisis de corte transversal identificando variables independientes correlacionadas con las variables dependientes seleccionadas.
2	Análisis longitudinal acerca de la relación entre el problema abordado por el programa y la intervención sin uso de grupo de comparación, o bien usando grupo de tratamiento y de comparación, pero sin probar la comparabilidad entre ambos. Para los casos de programas universales se ubicarán en esta categoría aquellas evaluaciones que usen una comparación “antes-después”.
3	Estudio basado en un diseño cuasi experimental con grupos de tratamiento y de comparación comparables.
4	Estudio basado en un diseño cuasi experimental, con grupos comparables de tratamiento y comparación, que controla por otros factores, aparte de la intervención.
5	Estudio que aplica un diseño experimental.

Fuente: basado en Sherman ...[et al] (1998).

Los informes de evaluación de los programas evaluados a través de la EPG han sido escrutados con la finalidad de constatar la extensión y profundidad del escrutinio, la información recogida y si el informe captura los aspectos sustantivos y complejidades del programa evaluado.

La práctica chilena de evaluación de programas

Sobre el Procedimiento de Evaluación

La DIPRES aparece como el órgano central en las decisiones de qué programa evaluar y en la implementación misma del procedimiento. Las entrevistas dan cuenta que la decisión de evaluación de un determinado programa es exógena a este e, incluso, al órgano público del que depende. Un entrevistado explica que “*se le comunicó al Subsecretario que este programa iba a ser evaluado por la DIPRES*”. Otro entrevistado señala “*que es una confluencia liderada por la DIPRES*”. Otros dos entrevistados explican que la decisión de evaluar el programa “*surge de un compromiso con el parlamento*” y “*de la discusión presupuestaria, supongo; [porque] la DIPRES acordaba con el Congreso cuántos programas entraban en cada proceso de evaluación*”. También se señala que “*el interesado en hacer esa evaluación es un tercero; [por un lado] la DIPRES y, por otro, el financista*”.

Aun cuando administradores de programas y autoridades de organismos públicos asumen una discusión y acuerdo entre el Congreso y la DIPRES en la determinación de los programas a evaluar, la visión que expone un ex parlamentario, en el siguiente comentario, muestra que esa decisión se centra casi completamente en la DIPRES: “*Los [programas los] asignaba siempre la DIPRES. O sea, se suponía que éramos los parlamentarios, pero la DIPRES nos proponía a los parlamentarios estos programas [los que serían evaluados]. Entonces, a lo más, uno podía meter un programa o dos programas, o sea, decir: ‘oye, ¿para qué vas a evaluar este? Nos interesa evaluar este [otro programa]’. Normalmente uno metía un programa que uno detectaba que no estaba andando bien en la lista -y que la DIPRES no quería meter, porque era un programa políticamente muy importante y si no estaba andando bien, era un problema-, pero de los 20 programas que se evaluaban cada año, los parlamentarios definíamos dos, y los otros 18 programas los proponía la DIPRES*”.

Por otro lado, programas y organismos se asignan un rol poco significativo en la designación de los evaluadores. Un entrevistado comenta que “*no tuvimos ningún rol [en la designación de los evaluadores], los identificó la DIPRES*”. Otro señala que “*hubo un abanico de opciones, y recuerdo que nosotros recomendamos a alguno de los panelistas, pero a modo informativo*”. Un tercer entrevistado explica que “*la DIPRES nos sugirió evaluadores y nosotros teníamos la posibilidad de objetar los evaluadores, para evitar que se produjera doble involucramiento entre entidades ejecutoras del programa vs. quienes estaban evaluando*”.

La relación que se establece entre evaluadores y evaluados, aunque sea conflictiva, es esencial, no tan solo para disponer de información relevante para la evaluación misma, sino también porque se

constituye en una instancia de reflexión y clarificación acerca del diseño del programa y su implementación. Un entrevistado señala que *“es muy importante que haya una implicación emocional y política de los evaluados en la evaluación”*. Otro entrevistado agrega que *“el hecho mismo de reflexionar sobre la cadena de impacto o resultados es útil para el ejercicio presente [del programa]”*. Un testimonio indica que *“hubo un trabajo intenso por parte del programa, más que nada para explicar en qué consiste este programa, no solamente al panel de expertos, [sino] que también a la gente de la DIPRES; porque hay visiones, de repente, parciales [respecto de] en qué consiste este programa, cuál es la población objetivo que atiende, qué se pretende [con él], bajo qué normas se rige”*. Otro entrevistado explica que *“había un flujo de información bien interesante, porque los profesionales entregaban información y el panel, de alguna manera, la retrucaba; entonces había un diálogo entre expertos (...) Recuerdo un par de detalles en los cuales el panel insistió en emitir juicios que a los profesionales de acá no les parecía que era una buena interpretación, y que mantuvieron la observación. Pero sí había un flujo de información”*. Otro entrevistado explica que *“nuestra participación en la evaluación fue bastante activa. El rol de la DIPRES, en algún momento, fue más administrativo, diría yo, y en algunos casos (...) zanjó ciertas situaciones en que entrábamos en conflicto con los evaluadores (...) nuestro gran problema con la consultora seleccionada [para hacer la evaluación] fue que había participado en el diseño original del programa; por lo que su visión era sesgada, según nuestro punto de vista”*.

El rol del impulsor de la evaluación es complejo y crítico a la vez, porque, por un lado, debe asegurarse que el proceso de la evaluación transcurra y que la información necesaria esté disponible; pero, por otro lado, ello puede afectar el clima de entendimiento y cooperación que debe darse entre evaluador y evaluado. Un testimonio señala que *“finalmente, el que impone los términos técnicos de la evaluación, los parámetros, la finalidad, los recursos, [es la DIPRES]; entonces se genera una triangulación, incluso incómoda para los evaluadores, porque ellos están acostumbrados a trabajar para y en beneficio del que está siendo evaluado; y en esta situación se los pone como una especie de ‘contralor’, entonces se pierde esa relación de colaboración, de irse nutriendo entre el evaluador y el evaluado, porque se genera, más bien, una relación de Contraloría y no fluye, no ha fluido para nada”*. Otro testimonio señala que *“nosotros no elegimos ni los indicadores ni el procedimiento”*.

Sobre la receptividad de la DIPRES a las objeciones de los evaluados un testimonio apunta lo siguiente: *“el gran problema con la consultora seleccionada fue que participó en el diseño original del programa, por lo tanto su visión era sesgada según nuestro punto de vista (...) Se lo advertimos [a la DIPRES] en la etapa de postulación, pero la DIPRES pensó que no era invalidante (...) ellos tenían el diseño original del programa en su cabeza y nosotros éramos los que lo habían ejecutado, habíamos ajustado ese diseño; entonces era una competencia; para nosotros era como pelear el padre con el hijo, era una pelea perdida y hasta el final fue eso (...) sentíamos que no estábamos frente a un evaluador que fuera imparcial, porque eso es lo que más necesita uno, alguien que esté dispuesto a escucharte súper abierto y aquí ellos querían evaluarnos con el modelo del programa que ellos tenían en la cabeza”*.

Lo que los comentarios reseñados muestran es que la DIPRES no solo impulsa la evaluación de programas públicos, sino que es ella quien -casi sin contrapeso- identifica los programas que serán evaluados, aporta los términos técnicos, los parámetros, la finalidad y los recursos. Así, entonces, la evaluación de programas en Chile es una práctica que se ha ido implementando y expandiendo bajo la orientación de la DIPRES. Ello hace que, en esta práctica, los evaluadores o paneles de evaluación no sean, en estricto rigor, independientes, sino que están sujetos a las directrices de quien los mandata para tal cometido, cuya función principal se orienta al control presupuestario.

Ello tiene, al menos, dos connotaciones. Por un lado, la de la existencia de un organismo público que impulsa y materializa prácticas evaluativas en el sector público chileno y con ello expande la cultura de la evaluación, siguiendo la *“larga tradición chilena de preocupación por el adecuado uso de*

los recursos públicos” (Olavarría y Figueroa, 2010: 36). Por otro, el hecho que el organismo que impulsa las prácticas evaluativas sea quien asigna los recursos y quien tutela el proceso es visto como una situación que resta objetividad e independencia a la evaluación, pues *“la DIPRES es juez y parte”* (entrevista a ex parlamentario), con *“una preocupación centrada en el control del gasto”* (entrevista a ex administrador de programa y ex autoridad pública).

La crítica que esboza esta segunda connotación es expresada con dureza por uno de nuestros entrevistados, quien ha cumplido diversos roles en el proceso de la política pública: *“la DIPRES es un poder suprajerrárquico. Hoy día, la DIPRES es, prácticamente, el gran gestor de la política pública en todos los ámbitos y no tiene contrapeso. Ellos saben tanto de bioquímica, como de ingeniería mecánica, como de política de desarrollo urbano y, además, de prospección financiera. Entonces, ellos saben de todo. Tienen especialistas en todo. Y cuando uno se los discute, ellos llegan a la conclusión de que uno es conflictivo. Cuando tú te logras sentar frente a ellos con una cuota de poder equivalente, puedes cambiar el dial. O sea, yo te diría que, en mi experiencia, cuando vino un interlocutor con componentes políticos y autoridad política, pudieron producir cambios que llevaron a que la política pública pudiera cumplir su objetivo, relativamente. Cuando no existe aquello o dejó de existir aquello, la burocracia impuso sus reglas y muchas veces impidió que la política pública pudiera lograrse”*.

Sobre la “usabilidad” de la información provista por las evaluaciones

En general, los entrevistados que han cumplido roles de administradores de programas o autoridad pública manifiestan que recibieron los informes, que estos fueron conocidos y discutidos en el organismo. Un testimonio señala que *“de los programas principales, sí nos llegaban los informes”*. Un entrevistado agrega que *“lo lógico sería que las autoridades reflexionaran con sus equipos, cosa que se producía, porque eran temas muy sensibles”*. Otro comentario expresa que *“hubo que hacer un análisis a las inferencias y a los comentarios que venían en los preinformes y en el informe final; entonces hubo una dedicación de nosotros de analizarlos para hacer una discusión, que haya una consistencia y una contraposición en aquellos aspectos que no eran compartidos”*.

Para los organismos y programas ha sido importante responder a los informes de evaluación, porque los evaluadores se *“pueden equivocar o no tener la información adecuada y porque hay una visión desde otra perspectiva”* (entrevista a administrador de programa).

Asimismo, el abanico de testimonios muestra reacciones favorables y críticas con los informes de evaluación. Por un lado, opiniones que señalan que *“la evaluación que hizo la DIPRES, a mí me parece bastante concordante con la realidad del programa”* (entrevista a administrador de programa). Otro administrador de programa agrega: *“fue un trabajo agradable, donde se expusieron los puntos de vista del programa y donde, entre comillas, podríamos decir que [el informe final] fue de mutuo acuerdo, porque llegamos finalmente a un documento donde se establecen los compromisos y donde definimos las fechas de los compromisos y el monitoreo en forma conjunta”*. Por otro lado, un testimonio señala que, aunque advirtieron de los errores en la evaluación, fue *“una pelea perdida (...) nosotros dijimos ‘te equivocaste en el sujeto [entrevistado], no es el que sabe’; [además en la encuesta] el ‘n’ era bajo, no era representativo, pero ya estaba hecho, ya no había vuelta atrás”*. Una ex autoridad pública entrevistada agrega: *“cuando llegaban los informes finales de parte de los gerentes, había algún grado de decir: ‘sí, pero [esto] no es perfecto’; pero, en general, eran bien recibidos, eran recepcionados positivamente”*.

Luego de finalizada la evaluación surgen los compromisos entre la DIPRES y el programa, cuestión que normalmente se deriva de las recomendaciones contenidas en la propia evaluación. La DIPRES luego hace un seguimiento al cumplimiento de los compromisos, lo que se extiende por el tiempo declarado en que los compromisos deben cumplirse. El Cuadro 4 muestra la cantidad de compromisos surgidos de las evaluaciones a que fueron sometidos los programas de la muestra y el grado de cumplimiento de ellos al 30 de junio de 2011.

Cuadro 4
Compromisos y cumplimientos de los programas de la muestra

Nombre del Programa	N° Compromisos	N° Cumplidos	N° Parcialmente cumplidos
Programa de Administración Directa (Panel 2000)	8	8	
Programa de Administración Directa (Panel 2007)	30	26	2
Programa Orígenes (Panel 2004)	No se encuentran compromisos en página web de la DIPRES		
Programa Orígenes (Panel 2010)	6	1	4
Programa de Prevención y Control del VIH/SIDA y Enfermedades de Transmisión Sexual (ETS) (Panel 2000)	9	8	
Programa de Prevención y Control del VIH/SIDA y Enfermedades de Transmisión Sexual (ETS) (Panel 2010)	5	2	2
Chile Barrio (Panel 2002)	25	23	
Seguridad y Participación (Panel 2010)	3	1	1
Plan Cuadrante (Panel 2007)	No efectúa compromisos		
Becas Presidente de la República (Enseñanza Media) (Panel 1999)	4	4	
Programa Administración de Concesiones (Panel 2007)	9	3	4
Programa de Mejoramiento de la Calidad y Equidad de la Educación Superior (MECESUP) (Panel 2004)	19	12	
Programa de Mejoramiento Urbano y Equipamiento Comunal (PMU) (Panel 1999)	4	4	
Programa de Mejoramiento Urbano y Equipamiento Comunal (PMU) (Panel 2011)	No se encuentran compromisos en página web de la DIPRES		
Programa para la Recuperación de Suelos Degradados (Panel 2000)	9	9	
Programa para la Recuperación de Suelos Degradados (Panel 2009)	11	8	2
Programa Red Tecnológica Educacional - Enlaces (EI 2003)	66	3	
Fondo de Tierras y Aguas Indígenas (EI 2001)	63	60	
Programa Nacional de Mujeres Jefas de Hogar (EI 2009)	13	3	8
Programa de Mejoramiento de la Gestión (EI 2007)	19	19	
Programas de Empleo con Apoyo Fiscal (EI 2004)	12	12	
Bonificación en la Contratación de Mano de Obra (EI 2008)	11	9	2
Programa de Educación y Capacitación Permanente, Chilecalifica (EI 2007)	No se encuentran compromisos en página web de la DIPRES		
Programa Chile Solidario (EI 2010)	No se encuentran compromisos en página web de la DIPRES		
Programa de Mejoramiento Urbano (EI 2002)	39	35	
Beca Presidente de la República (EI 2006)	8	6	
Programa Chile Barrio (EI 2006)	31	31	
Programas de Fomento CORFO (EI 2009)	No se encuentran compromisos en página web de la DIPRES		

Fuente: elaboración propia.

El Cuadro 4 da cuenta de un alto grado de cumplimiento de los compromisos -salvo el programa Enlaces-, lo que sugiere que los informes de evaluación han tenido importantes efectos en la gestión de los programas. Esta visión se contrapone, sin embargo, con el testimonio -invariable- aportado por los entrevistados. Consultados si la evaluación aportó información novedosa, un administrador de programas responde que “*fue ratificar, básicamente [algo que se sabía]*”; otro administrador de programa acota que “*no fue un insumo, en estricto rigor*”; otro administrador de programa agrega que los informes “*no tuvieron ninguna utilidad para orientar el programa [o] para hacer cambios en la gestión*”. Un testimonio adicional señala: “*Yo te diría que, en general, [los administradores de los programas] saben, saben todo lo que va a pasar [en la evaluación]*”.

Los entrevistados que han ejercido roles de autoridad pública apuntan en la misma dirección respecto de la misma pregunta. Una ex autoridad pública, que también desempeñó roles de administrador de programa, señala: *“Yo tengo la sensación de que, en general, nunca hubo una evaluación que a mí me sorprendiera (...), pero yo pienso que eso tiene que ver con el nivel de información que uno maneja”*. Otra ex autoridad pública expresa que *“los administradores de los programas tienen una impresión de cómo se están haciendo las cosas, son los dueños de su negocio (...) [me] iban informando y teníamos conversaciones permanentes respecto de cómo iban las cosas; o sea, no es que llegara una información que uno no sospechara o no tuviera algún indicio”*. Una tercera ex autoridad pública anota: *“tanto a nivel de impacto de la cobertura que tuvo el programa como de los resultados esperados, hubo una coincidencia entre lo que nosotros habíamos proyectado y los resultados de la evaluación”*. Una cuarta ex autoridad expresa que *“los paneles rara vez te decían algo distinto de lo que tú esperabas que te dijeran”*. Otra ex autoridad agrega: *“no le presté gran interés, no fue trascendente [el informe de evaluación]”*.

Desde la experiencia del Congreso, se anota que la evaluación *“está llena de la visión que uno tenía de antes; los paneles confirmaban el sesgo con el cual se pedía la evaluación”*. También se señala que *“la gran mayoría de los parlamentarios ni siquiera conocían la existencia de una evaluación. El debate se produce en un grupo muy reducido que es la Comisión Mixta de Presupuesto y, aún más, en la Subcomisión Mixta de Presupuesto, que son cinco: tres diputados y dos senadores, que abordan la discusión del presupuesto con relativa profundidad. Tres sesiones de cuatro horas para discutir el presupuesto de un ministerio, no necesariamente uno puede decir que hay una discusión en profundidad. En ello, algunos parlamentarios consultan los instrumentos de evaluación, otros ni siquiera los consultan. O sea, la verdad es que la evaluación de la política pública no es una variable a considerar en la discusión presupuestaria”*.

Un tema que deja ver la cita recién expuesta es la falta de capacidad del Congreso para procesar la información que provee la evaluación. La cuestión también es resaltada por la Presidenta de la Comisión Mixta de Presupuesto en la discusión de la Ley de Presupuesto 2010, quien, con motivo de la discusión sobre programas de deporte, expresa: *“La verdad es que estamos pidiendo una cantidad de información que yo no sé quién va a procesar ni quién la leerá, y que además significa una buena cantidad de funcionarios dedicados a esa labor (...). Lo que pasa es que nosotros debíamos tener realmente un equipo fuerte de seguimiento, pero, sobre todo, de evaluación de impacto, porque, en definitiva, el saber en qué se gastó es bastante menos importante que saber que si lo que se gastó significó finalmente una diferencia en la vida de la gente”* (Congreso Nacional de Chile, 2010: 1132-1133).

Asimismo, la revisión de las historias de las leyes de presupuesto del 2006 al 2011 muestra escasas referencias⁶ a evaluaciones de programas, y los informes de evaluación no son una fuente de consulta rutinaria ni recurrente en los debates y decisiones parlamentarias acerca de las asignaciones presupuestarias a los programas. La revisión también muestra que solo muy ocasionalmente, en casos de abierta disputa política, hay referencias superficiales y generales a aspectos parciales de las evaluaciones, y que el Congreso incluye programas a ser evaluados, mediante protocolos de acuerdo con la DIPRES, cuando hay dudas de su efectividad, de probidad en su gestión o hay controversia política respecto de los propósitos que sirve.

Un claro ejemplo de lo anterior es el Programa Orígenes, que recoge escasas, pero duras críticas en la discusión de la Ley de Presupuesto de 2009 (Congreso Nacional de Chile, 2009) y cuya evaluación de impacto es acordada en la discusión de la Ley de Presupuesto de 2010 (Congreso Nacional de Chile, 2010: 1223 y 1250). Un hecho que llama la atención es que luego de haberse acordado la evaluación de impacto de este programa, en la discusión de la Ley de Presupuesto siguiente -la de 2011- el Congreso es informado que el programa sería discontinuado y que no habría fase tres (Congreso Nacional de Chile, 2011: 581-582). Ello hace que los resultados de la evaluación de impacto

del programa sean extemporáneos y de ninguna utilidad para el diseño y gestión del programa y el organismo que los alberga, pues la decisión de terminarlo había sido tomada antes que esa información estuviera disponible.

Asimismo, la revisión de los Balances de Gestión Integral da cuenta de referencias a los aspectos formales de las evaluaciones, de los compromisos contraídos con DIPRES y del grado de cumplimiento de ellos, pero no hacen referencia a hallazgos de los informes o de implicancias para el programa, más allá de los aspectos formales ya comentados.

Junto con mostrar que las evaluaciones no aportaban información novedosa, los testimonios indican que ellas tampoco han tenido incidencia destacable en la marcha de los programas. Testimonios de administradores de programas señalan que *“para la gestión del programa no fue útil”*; que no generó cambios en la gestión, ni en las prioridades, ni en la asignación presupuestaria, ni en las decisiones políticas, ni a nivel institucional; que *“en realidad, el informe no tuvo ninguna utilidad”*. Opiniones de autoridades expresan que *“nosotros participamos, simplemente porque había que participar, no le dimos mucha importancia (...) lo hicimos como especie de saludo a la bandera del control de la DIPRES sobre las políticas públicas”*; que *“algunos cambios se hicieron (...) cuando eran cambios puntuales, muy específicos, tú puedes reorientar cosas muy concretas; pero hacer análisis globales, reorientar el programa en su conjunto, no”*.

El contrapunto que surge es interesante: mientras los informes de la DIPRES muestran la aplicación de un procedimiento de evaluación estructurado, que arroja resultados, que aporta información y que tiene efectos en la marcha de los programas, cuestión que se expresa en los compromisos que se formalizan entre la DIPRES y el programa, y en el seguimiento de su cumplimiento, las opiniones expresadas por administradores de programas, autoridades y parlamentarios, y la revisión de la historia de la ley de presupuesto y los Balances de Gestión Integral muestran una realidad opuesta.

¿Qué explica, entonces, esta disparidad? Opiniones de administradores de programas señalan que *“varias de las recomendaciones no se pueden implementar”*; que *“muchas veces los evaluadores no tienen esta dimensión de la intersectorialidad [del programa] (...) [y no comprenden que] nosotros no somos ejecutores, somos coordinadores”*; y que *“hay que darles más tiempo y buscar instrumentales que den cuenta de la complejidad [de los programas], porque muchas veces se evalúan más del punto de vista económico, pero hay otros elementos que se escapan (...) y hay que tener otras miradas [porque] son programas que están afectando a las personas”*. También se señala que: *“muchas veces las decisiones que se toman son menos técnicas de lo que uno quisiera (...) porque las evaluaciones del panel de expertos tampoco tienen un peso tan grande, como uno dijera ‘aquí tengo la evaluación del panel de expertos, háganme caso’. Tampoco tienen un impacto tan grande desde el punto de vista político. Fíjate que tiene más impacto lo que pueda decir la Contraloría respecto de las auditorías que hace, tienen más impacto [aunque] son menos técnicas (...). Entonces, [las evaluaciones] no tienen el peso suficiente como para que realmente la gestión que uno realiza dé un giro, pero lo que me dice la Contraloría sí, con razón o sin razón”*.

Otro testimonio señala que es necesario que la evaluación logre capturar la especificidad del programa, que adecue la *“metodología conforme se vaya incrementando la complejidad del programa, y por complejidad me refiero a la multiplicidad de objetivos que tenga y a la multiplicidad de actores que están involucrados”*.

Opiniones de autoridades coincidentemente señalan que *“el procedimiento era demasiado rígido, demasiado formalista, rutinario, no había una evaluación cualitativa, era muy basada en indicadores; era más bien un filtro de la DIPRES, [en el] que nosotros no habíamos contribuido ni en la elección de los indicadores, ni en el procedimiento”*; que *“[la evaluación] estaba mal hecha, esa era la percepción nuestra”*; que *“hay que entender también los momentos de los gobiernos (...); en los años electorales y de salida no alcanzas a hacer grandes reorientaciones, porque no tiene sentido; además eso implica*

cambios presupuestarios y los mismos tiempos presupuestarios no te alcanzan a cubrir esos cambios”.

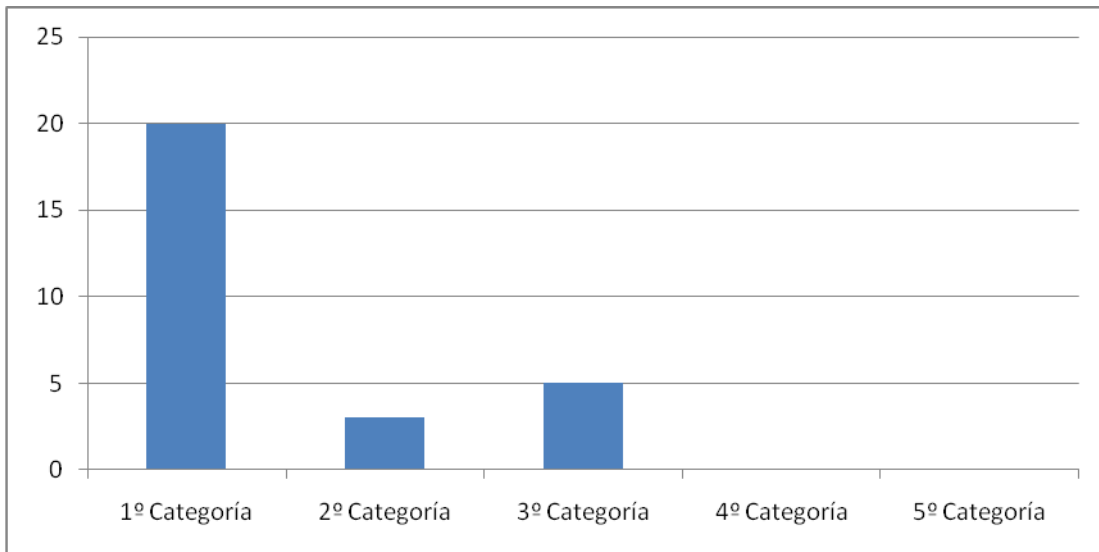
Un ex parlamentario entrevistado señala que las evaluaciones tienen escasa incidencia en las decisiones del Congreso, porque aportan poca información relevante para la toma de decisiones proactivas: *“Yo tengo la impresión de que los informes de las evaluaciones tienen poca prospectiva; o sea, son como verdaderas autopsias. Pero hacer una autopsia de política pública es solo radicarse cuando el muerto ya está muerto, pero no identifica las causas ni propone, muchas veces, soluciones”.* También se señala que lo que verdaderamente influye es *“la percepción de realidad, la información que los parlamentarios recogen en terreno, en sus distritos, de la ciudadanía que se comunica con ellos, que les expresa sus inquietudes, sus molestias; los lobbies respectivos, tanto de instituciones públicas como privadas con respecto a cada área de política pública. Eso es lo que más influye. No hay un instrumento de evaluación de política pública que los parlamentarios recojan. Por el contrario, yo te diría que en los últimos dos presupuestos, el análisis más en profundidad lo hacen equipos técnicos externos que se han formado del Instituto Libertad, del Jaime Guzmán, Igualdad, CIEPLAN, que analizan mucho más en profundidad el presupuesto que los propios parlamentarios. Ellos van diciendo: ojo aquí, ojo allá, pidamos toda esta información”.*

Visto en perspectiva, lo que la evidencia muestra es que hay una extendida percepción, entre administradores de programas, autoridades públicas y parlamentarios -al menos, los entrevistados-, que los modelos de EPG y EI que se han implementado no logran capturar los aspectos sustantivos de los programas, lo que hace que las decisiones que los poderes públicos toman sobre ellos sean recurrentemente políticas y con poca referencia a los informes de evaluación. De este modo, los resultados de las evaluaciones, los compromisos y su cumplimiento son vistos, por los programas y organismos, como cuestiones formales que hay que cumplir, pero con pocos efectos prácticos sobre la gestión de los programas.

Sobre el rigor de las evaluaciones

El análisis de los informes de EPG y de EI incluidos en la muestra expone los principales problemas metodológicos detectados. Los 23 programas de la muestra acumulan 31 evaluaciones. De ellas, 18 son de la modalidad EPG, 10 de EI, y tres evaluaciones en profundidad.

Gráfico 1
Distribución de las evaluaciones de la muestra
según la Escala de Rigor Metodológico



Fuente: elaboración propia.

Nota: Este gráfico considera la asignación de la categoría de rigor metodológico a cada evaluación de impacto. A ello se le han agregado las evaluaciones del tipo EPG, las que han sido asignadas a la categoría 1. Debe considerarse que hay tres evaluaciones en profundidad que no han sido incluidas en el gráfico.

Las evaluaciones de impacto de los programas incluidos en la muestra logran alcanzar bajas categorías en la Escala de Rigor Metodológico. De los 13 programas que el sitio web de la DIPRES incluye en el módulo de EI, tres corresponden a evaluación en profundidad, por lo que no es posible aplicarles la escala de rigor metodológico, cinco se ubican en la categoría 3 de la escala, tres alcanzan la categoría 2, en tanto que dos evaluaciones solo llegan a la categoría 1, donde 1 es el más bajo rigor y 5 el máximo rigor. Ello muestra debilidades de los resultados y de la validez de las conclusiones de las evaluaciones analizadas.

El análisis muestra que las principales debilidades de las evaluaciones de impacto revisadas se ubican en los siguientes aspectos:

- Falta de una línea de base al momento de inicio de la implementación del programa. Con frecuencia, la línea de base es establecida con posterioridad a la puesta en práctica de la intervención, lo que implica que la muestra ha sido contaminada y, con ello, que la estimación del impacto real se hace improbable.

- Líneas de base construidas por autorreportes basados en recordación de los encuestados respecto de cómo era su situación al inicio del programa o si este solucionó, en algo, el problema objeto de la intervención. Ante la falta de una línea de base, con frecuencia las evaluaciones acuden a esta opción de línea de base, pero ella genera sesgos en las estimaciones.

- Falta de control y de acciones remediales frente al desgaste de la muestra. Aunque de la lectura de los informes se deduce un desgaste de la muestra, no hay constancia en ellos que haya habido identificación de los sesgos que esta situación produce y acciones que permitan abordarlos para generar estimaciones insesgadas.

- Comparaciones antes-después. Aunque esta es una técnica usada en evaluaciones de impacto de programas de cobertura total, se observa su aplicación en evaluaciones de programas de cobertura

parcial y falta de análisis de variables contextuales que pudieran estar incidiendo en los resultados observados.

- Construcción de grupos de comparación por manipulación estadística, sin inclusión de pares de mayor similitud. Aunque este es un problema de menor envergadura, comparado a los anteriores, no deja de tener efectos sobre los potenciales sesgos de las estimaciones. La técnica usada para la construcción del grupo de comparación es el “*propensity score matching*”, pero frente a la dificultad de identificar el “par” que presenta la más alta puntuación se ha usado la opción del “vecino más cercano”, lo que implica una mayor diferencia respecto de la población intervenida. En todo caso, insistir con el “par más parecido” implica dedicar mayor tiempo y un mayor costo.

Por otro lado, la revisión de las evaluaciones tipo EPG⁷ muestra, en general, convergencia en los resultados de evaluaciones hechas en períodos diferentes sobre un mismo programa⁸, aunque no hay un examen detenido acerca de las conclusiones del panel anterior. De los programas de la muestra que presentan más de una evaluación, solo dos presentan una revisión de las conclusiones de la evaluación previa, los compromisos adquiridos y el grado de cumplimiento de ellos. En esos casos, los evaluadores advierten que el cumplimiento de las recomendaciones es formal y que los compromisos no presentan un cumplimiento sustantivo.

El análisis de este tipo evaluaciones deja ver, sin embargo, dos situaciones que inciden en las conclusiones a que arriban y, en consecuencia, en los efectos que estas evaluaciones podrían generar sobre la marcha de los programas.

Las evaluaciones EPG toman como insumo básico del trabajo el marco lógico del programa. Este, a su vez, recurrentemente asume en la declaración de su finalidad y propósitos que es el programa, por sí mismo, el encargado de cumplirlos. Sin embargo, la realidad de la gestión muestra que la labor de ellos tiende a centrarse, más bien, en la articulación de la llamada “oferta pública” y en hacer que la demanda por servicios existentes en la comunidad encuentre o se contacte con esa “oferta pública”. De este modo, entonces, variados programas son implementados en los hechos con un modelo de gestión de red, pero sus marcos lógicos y, consecuentemente, su evaluación los asume como una unidad en sí misma, con un modelo tradicional de gestión arriba-abajo.

También es posible observar que, con frecuencia, el marco conceptual y las hipótesis de impacto o resultados de la intervención son poco explicitados, tanto en la descripción del programa, en su marco lógico y en la evaluación. Ello le resta claridad al proceso de implementación del programa, generando una propensión a resaltar las cuestiones administrativas, cuestión de la que es difícil abstraerse para quienes están en la implementación diaria de la intervención.

Sobre el tema, un testimonio apunta lo siguiente: *“El sistema de la matriz de marco lógico como elemento de planificación se ve teñido por la lógica de poder indicar que son de productos. En general, muy pocos marcos lógicos son de indicadores de resultados: mucho porcentaje de satisfacción, con tasas declaradas que son bajas, pero los indicadores de satisfacción no tienen mucha variabilidad. No hay mucha reflexión sobre cómo se fijan las metas, son cosas muy intuitivas. En general, hay malos marcos lógicos, desde lo básico de cómo están definidos. La principal deficiencia en la planificación es que no logran parametrizar cuáles son los cambios finales que se esperan, los resultados. Entonces, finalmente son casi indicadores de actividades”*.

Aunque estas son cuestiones que tienen directa relación con el diseño de los programas, también tienen considerables efectos sobre su evaluación y la constatación de cumplimiento de objetivos, obtención de resultados y verificación de impactos.

Visto en perspectiva, la revisión de informes de evaluación de programas de esta muestra intencionada expone dificultades en la identificación de impactos, confirmación de resultados y efectividad en la implementación del programa. Una cuestión que será conveniente dilucidar es si ello está relacionado con el bajo efecto real que los informes de evaluación han alcanzado en las decisiones de los poderes públicos o de administradores de programas y organismos públicos que los albergan,

según reportan las entrevistas efectuadas y la revisión documental de la historia de la ley de presupuesto y Balances de Gestión Integral.

Discusión

El análisis muestra un contrapunto interesante. Mientras la DIPRES exhibe que los procesos de evaluación se cumplen y que ello genera efectos en los compromisos que se adquieren con la administración de los programas y los organismos públicos que los albergan, tanto las entrevistas a parlamentarios, administradores de programas y autoridades, como la revisión documental muestra que las evaluaciones tienen poca incidencia en las decisiones que se toman sobre los programas. Esta constatación es coincidente con el informe de evaluación del Programa de Evaluación de Programas del Banco Mundial, que siete años antes de este estudio señalaba que “los congresistas (...) no parecen utilizar esta información ni aportar creativamente al PE, [aunque] están posicionados para hacerlo” (Banco Mundial, 2005: 11).

Las razones que podrían estar incidiendo en esta situación, de acuerdo a la evidencia recogida en el presente trabajo, serían:

1. Las decisiones sobre los programas son políticas y en ello las evaluaciones no son el insumo central. Como manifiesta un administrador de programa entrevistado, respecto que las evaluaciones tienen “*una utilidad que es netamente técnica, pero este tipo de decisiones no son solamente técnicas*”; o como señala otro entrevistado: “*es interesante el informe del panel, pero no tiene la fuerza*”.

2. En la discusión sobre los programas, los parlamentarios acuden a información proveniente de sus distritos, asesores, *lobby* de organismos públicos y privados, y de los institutos parapartidarios, antes que de las evaluaciones efectuadas a los programas. En ello puede estar incidiendo la tensión que se advierte entre los parlamentarios y la DIPRES, a quien estos consideran que es “*juez y parte*” y que “*es un poder suprajerárquico*” del propio Poder Ejecutivo.

3. Los administradores de programas y autoridades expresan que los informes de evaluaciones no les aportaron información novedosa, que las evaluaciones son algo rígidas, rutinarias y formalistas, y que no recogen los aspectos sustantivos de los programas y sus complejidades. También se advierte tensión entre administradores y autoridades, por un lado, y la DIPRES, por otro. Esta tensión la expresa con vehemencia una ex autoridad entrevistada: “*la injerencia de la DIPRES en evaluar políticas públicas desde un supraministerio se hizo terrible (...). Basta con que te evalúen mal una política, con el instrumento que ellos mismos diseñan, cuyo objetivo tú ni siquiera conoces, los indicadores no los discutieron contigo, te ponen mala nota y te friegan la política pública que es de tu ministerio. El riesgo político lo corres tú (...); al señor Ministro de Hacienda le importaba un ‘pucho’, pero la huelga me la comía entera yo*”.

4. Los problemas de rigor metodológico identificados en los informes de evaluaciones, lleva a que no haya certeza que los resultados y estimaciones de impacto sean insesgados.

5. El resultado de la evaluación también está asociado a la actuación de la contraparte. Ello hace necesario que quien actúe de contraparte de los evaluadores, por parte del programa evaluado, desempeñe un rol activo, con capacidad de interlocución técnica y de representatividad del organismo y del programa. Un ex administrador de programa y ex autoridad explica: “*hay que tener muchísimo cuidado en cómo se va dando institucionalmente [la evaluación] y cómo se dan garantías a la contraparte. En este caso particular, yo estaba al otro lado y algo sabía de evaluación, por lo tanto podía ser, digámoslo así, un evaluado activo, y podía cuestionar la evaluación. Pero en muchos casos es muy probable que haya habido evaluados pasivos, sin la capacidad técnica como para cuestionar lo que hacía el evaluador, porque estaban en una situación de gran asimetría y pueden haber ocurrido casos en los cuales programas fueron sencillamente mal evaluados y pueden haber sido programas que hayan estado haciendo una diferencia y hayan sido cortados, y eso es un riesgo que no debiera ocurrir*”. Otro testimonio agrega que es necesario que la contraparte de la evaluación, por parte del

programa, tenga una visión *“tanto política, como de política pública y técnica (...) que tenga cierta jerarquía en la institución, que tenga un dominio técnico importante, aunque no tan detallado, que tenga un conocimiento del funcionamiento de la institución y entienda cuál es el sentido de política pública, incluso el sentido político del estudio”*.

Aunque estas líneas muestran una visión crítica hacia la práctica chilena de evaluación de programas por parte de los entrevistados, de la revisión documental y del análisis de rigor metodológico, también es dable señalar que la evaluación de las intervenciones del Estado ha avanzado gracias al empuje y tesón que se ha puesto desde la DIPRES, con el ánimo de aportar información relevante a las decisiones de política pública y a su implementación. Dicho en otros términos, aun con las dificultades y críticas anotadas, el avance logrado en materia de evaluación de programas y la expansión de una cultura de la evaluación se debe, en gran medida, al esfuerzo de la DIPRES por aportar elementos técnicos y metodológicos a la gestión de los programas y al proceso de toma de decisiones de política pública en Chile.

La evidencia recolectada muestra, sin embargo, que un órgano evaluador de políticas públicas instalado en el Poder Ejecutivo -como la DIPRES- genera tensión en el proceso de formulación -en la decisión- y en la implementación de programas públicos, pues se la ve, por parte de los evaluados, como “juez y parte”, que desarrolla una evaluación más bien formal, rígida, que no aporta información novedosa y que no captura -al menos adecuadamente- los aspectos sustantivos y complejidades del programa. Aunque el argumento central que apoya la existencia de un órgano evaluador en el Poder Ejecutivo, como la DIPRES, es la “usabilidad” de la información para la toma de decisiones del gobierno y la asignación presupuestaria, la evidencia recogida muestra que autoridades sectoriales y administradores de programas acuden a información propia, y que para los parlamentarios los informes de evaluación no constituyen una información de alta relevancia en sus decisiones.

Asimismo, dado que la evidencia muestra que las evaluaciones -particularmente las EPG- no aportaron información novedosa para los administradores de programas y autoridades sectoriales -aunque se constituyeron en una instancia de reflexión acerca del devenir del programa-, surge inmediatamente la pregunta acerca de si resulta efectivo instalar las evaluaciones intermedias -de seguimiento y monitoreo- en un órgano del gobierno central, o si privilegiar el hecho que son prácticas propias de la gestión de programas que aportan información relevante a los encargados de la implementación.

Ello, indudablemente, hace surgir la discusión acerca de la organización de un sistema de evaluación de programas en Chile. La evidencia recogida insinúa algunas claves para ello:

- Dado que autoridades y administradores de programas recogen información periódica acerca del desempeño de las intervenciones, la descentralización de las evaluaciones de seguimiento y monitoreo, a nivel de ministerios, aparece como una medida consistente con la evidencia. Ello, a su vez, implicaría extender la cobertura de este tipo de evaluaciones, superando el número histórico que anualmente se implementa, y, con ello, no solo expandir la cultura de la evaluación, sino que, fundamentalmente, incrementar la efectividad de las intervenciones públicas.

- Dado que las evaluaciones de impacto y de resultados implican dar un juicio -objetivo y basado en evidencia- acerca del aporte global del programa a la solución, disminución o contención del problema que aborda, y dado que ello se da cuando el programa mismo o una etapa sensible de él ha concluido, parece aconsejable que esta práctica se institucionalice en un órgano ajeno al financiamiento, a la implementación, al interés político sobre el programa y a las decisiones que se han tomado sobre él. Así, el testimonio, respaldado por la evidencia recogida, que debe entregar este tipo de evaluación puede permitir a los poderes públicos arribar a decisiones con fundamentos ciertos frente a los actores políticos, institucionales y ciudadanos. Este tipo de práctica implica, asimismo, una evaluación a las evaluaciones efectuadas en las etapas anteriores del proceso de la política pública: de diseño, de eficiencia (costo beneficio y costo efectividad) e intermedia (seguimiento y monitoreo). Con

ello se genera una retroalimentación relevante para el permanente ajuste y perfeccionamiento del sistema de evaluación del país.

- Los temas relativos al rigor metodológico, marcos conceptuales de las intervenciones e hipótesis de impacto en el diseño de los programas hacen ver la necesidad que el órgano que impulsa las prácticas evaluativas desarrolle labores de los siguientes tipos:

- normalización metodológica, de modo que todos los actores del sistema de evaluación puedan disponer de herramientas que han sido analizadas, establecido su validez en la literatura especializada, su concordancia con los tipos de evaluación emprendidos y, finalmente, usables por los distintos actores;

- asegurar que los diseños de los programas se sustenten en marcos conceptuales basados en evidencia, nacional o internacional, validada tanto en rigurosos métodos de investigación social como en referencia a las características del problema que la intervención aborda;

- asegurar que el diseño de los programas exprese con claridad las hipótesis de impacto, de obtención de resultados, ya que este es un aspecto clave en las evaluaciones que se harán al programa;

- asegurar que antes de su aprobación y asignación de recursos, todo programa incluya un componente de evaluación rigurosamente construido, de modo de evitar falta de líneas de base al momento de la evaluación -o construir las mediante autorreporte de recordación o al tiempo después de iniciada la intervención- o carencia de información sustantiva, lo que puede sesgar los resultados de la evaluación.

Asimismo, la constitución de un sistema de evaluación, con prácticas descentralizadas a nivel de ministerios, hace necesario:

- La instalación de un nodo central que coordine, impulse y dialogue con los poderes públicos con la finalidad de consensuar el programa y calendario de evaluación, a la vez que informar de los resultados de las evaluaciones emprendidas.

- Generar permanentes actividades de formación, capacitación, difusión y extensión en las materias relacionadas con la evaluación de programas. Ello permitirá, por un lado, ampliar la masa crítica de evaluadores y, por otro, expandir la cultura de la evaluación y el interés ciudadano y político por ella. En este ámbito, aparte de las actividades tradicionales de capacitación y formación, las actividades de extensión, en las que se expongan a actores políticos relevantes y ciudadanos los resultados de la evaluación de un programa determinado -con réplica de los evaluados-, puede constituirse en una práctica que permita ampliar la cultura evaluativa en el país y acercar a actores políticos al uso de los resultados de las evaluaciones.

Visto en perspectiva, lo que la evidencia recolectada sugiere es la necesidad de acercar, articular y hacer sintonizar a los distintos actores -productores y usuarios de información- del sistema chileno de evaluación de programas.

Conclusiones e implicancias

El análisis de una muestra intencionada de programas sometidos a evaluación, tanto en la modalidad EPG como EI, muestra disparidad entre lo que expresa el proceso formal llevado a cabo bajo el impulso de la DIPRES, por un lado, y la visión que surge de las entrevistas a actores relevantes, la revisión documental y el análisis del rigor metodológico, por otro. Mientras el proceso formal expone un modelo de evaluación que aporta información relevante a la toma de decisiones, las entrevistas, revisión documental y análisis metodológico de las evaluaciones muestran una realidad diferente.

En el caso chileno, la práctica de evaluación de las intervenciones públicas ha sido impulsada y tutelada desde un órgano del gobierno central y asociada al ciclo presupuestario del sector público. La ventaja predicada de este modelo ha sido la “usabilidad” de la información que proviene de las evaluaciones para decisiones sobre los programas, su orientación, (re)diseño y gestión. La evidencia recogida muestra que autoridades ejecutivas sectoriales, administradores de programas y

parlamentarios acuden a información propia para las decisiones sobre los programas y su gestión, y que el cumplimiento del proceso es visto por ellos como una cuestión formal. De este modo, la “usabilidad” de la información queda circunscrita a las acciones que emprende la DIPRES en la continuidad de los programas y en la asignación presupuestaria, pero ello genera resistencia en autoridades sectoriales y administradores de programas respecto de ese órgano del gobierno central, porque es visto como “juez y parte”, que centra su decisión en cuestiones asociadas al control de costos, antes que en el reconocimiento a los aspectos sustantivos y complejidades de los programas.

Aun con este contrapunto y la crítica vehemente que surge de las entrevistas, dos cualidades positivas deberían, en justicia, resaltarse respecto de la actual práctica evaluativa chilena. Una, referida a que, aunque el proceso sea visto como formal, ha permitido instalar una práctica de evaluación de las intervenciones públicas, que ha ido expandiéndose, y que ha colaborado a que administradores de programas y autoridades sectoriales generen un seguimiento acerca del devenir del programa, sus resultados e impactos. Otra se refiere a que las evaluaciones practicadas han generado instancias de reflexión acerca de la marcha del programa para los propios cuerpos directivos, profesionales, autoridades y actores vinculados a él, cuestión que los entrevistados reconocen como positivo.

Las implicancias de la evidencia, del contrapunto y tensión que se da en torno a la práctica chilena de evaluación de programas, que recoge este estudio, sugiere que el diseño e instalación de la prometida Agencia de Evaluación de Políticas Públicas -que ha sido incluida en los programas de gobierno de la presidenta Bachelet y del presidente Piñera- haga que la práctica de la evaluación mantenga estabilidad en el tiempo, se asocie a instancias formales de decisión, incentive instancias de reflexión en los evaluados y genere información aceptada por los actores involucrados. En ello, el razonamiento expuesto sugiere desconcentrar las evaluaciones de seguimiento y monitoreo y, por otra parte, radicar en esa agencia las evaluaciones de resultados e impacto, porque daría mayor efectividad, agilidad y “usabilidad” de la información producida, a la vez que establecería una retroalimentación a las evaluaciones previas. Asimismo, la evidencia recogida expone la necesidad de superar la percepción de que quien lidera el proceso sea “juez y parte”, porque ello -de acuerdo a la evidencia recogida- augura una mejor aceptación de la práctica evaluativa y pronostica mayor “usabilidad” de los reportes.

Notas

¹ En ocasiones el panel también ha estado integrado por cuatro o cinco personas.

² Los informes se denominan “Análisis de Información Base para la Evaluación”, “Informe de Avance”, “Informe Preliminar Final”, “Informe Final” e “Informe Final Corregido”, y deben entregarse a fines de enero-febrero, marzo, abril, mayo y julio respectivamente (DIPRES, 2012).

³ Con recurrencia la literatura se refiere a este tipo de evaluación como “Evaluación de Procesos”. Cohen y Franco (2000: 110) señalan que ella “determina la medida en que los componentes de un proyecto contribuyen o son disfuncionales a los fines perseguidos, (que) se realiza durante la implementación y, por tanto, afecta a la organización y las operaciones, (que) busca detectar las dificultades que se dan en la programación, administración, control, para corregirlas oportunamente, disminuyendo los costos derivados de la ineficiencia, y (que) no es un balance final, sino una evaluación periódica”.

⁴ El cuasi-experimento es útil para programas de cobertura parcial. Ha habido casos en que en la EI se han incluido programas de cobertura total, por lo que evaluaciones del tipo de “controles reflexivos” han sido necesarias. Los “controles reflexivos” pueden ser del tipo “comparaciones antes-después” o “series de tiempo”. En ambos casos, la población intervenida es comparada consigo misma en dos momentos del tiempo a través de sucesivas mediciones.

⁵ La Evaluación de Programas Nuevos corresponde a esfuerzos desarrollados para que, a contar del 2009, los programas cuenten con línea de base y “que el diseño de la evaluación esté radicado fuera de la institución responsable del programa” (Arenas y Berner, 2010: 48).

⁶ La revisión de la historia de las leyes de presupuesto señaladas mostró que las evaluaciones de dos de los programas incluidos en la muestra fueron referidos en el debate acerca de la respectiva ley de presupuesto.

⁷ De haberseles aplicado la escala de rigor metodológico, las EPG habrían calificado en la categoría 1 (ver Gráfico 1).

⁸ Es de notar que esta apreciación es coincidente con el hecho que los tópicos abordados por los paneles son similares, cuestión asociada al hecho que es una evaluación estandarizada.

Bibliografía

- Arenas de Mesa, Alberto y Berner Herrera, Heidi (2010), “Presupuesto por resultados y la consolidación del sistema de evaluación y control de gestión del gobierno central”, Santiago, Ministerio de Hacienda. Dirección de Presupuestos. División de Control de Gestión.
- Banco Mundial (2005), “Chile. Estudio de evaluación de impacto del programa de evaluación de programas: evaluaciones de impacto y evaluaciones de los programas del gobierno”, Washington, Banco Mundial. Unidad de Reducción de la Pobreza y Gestión Económica para América Latina y el Caribe (Informe N° 34589-CL).
- Cohen, Ernesto y Franco, Rolando (2000), *Evaluación de proyectos sociales*, México, Siglo XXI Editores. 5. ed.
- Congreso Nacional de Chile (2009), Historia de la Ley N° 20.314, de Presupuesto Público del año 2009, Santiago, Biblioteca del Congreso Nacional de Chile.
- _____ (2010), Historia de la Ley N° 20.407, de Presupuesto Público del año 2010, Santiago, Biblioteca del Congreso Nacional de Chile.
- _____ (2011), Historia de la Ley N° 20.481, de Presupuesto Público del año 2011, Santiago, Biblioteca del Congreso Nacional de Chile.
- DIPRES (2012), Pauta para la elaboración de informe final de evaluación, Santiago, Ministerio de Hacienda. Dirección de Presupuestos. División de Control de Gestión, mimeo.
- Olavarría Gambi, Mauricio (2010), “Introducción”, en *¿Cómo se formulan las políticas públicas en Chile? Tomo 1. La modernización de la gestión pública*, Mauricio Olavarría Gambi (ed.), Santiago, Editorial Universitaria.
- Olavarría Gambi, Mauricio y Figueroa Huencho, Verónica (2010), *Una agencia de evaluación de políticas públicas para Chile: lecciones de la historia y de la experiencia internacional*, Santiago, Editorial Universitaria.
- ONU (1984), Pautas básicas para el diseño y uso de sistemas de seguimiento y evaluación de proyectos y programas de desarrollo rural en países en desarrollo, Roma, ONU. Grupo de Trabajo sobre Desarrollo Rural del Comité Administrativo de Coordinación, mimeo.
- Rossi, Peter y Freeman, Howard (1993), *Evaluation: a Systematic Approach*, Beverly Hills, Sage. 5. ed.
- Sherman, Lawrence; Gottfredson, Denise; Mackenzie, Doris; Eck, John; Reuter, Peter; y Bushway, Shawn (1998), *Preventing Crime: What Works, What Doesn't, What's Promising*, College Park, University of Maryland. Report to the United States Congress, prepared for the National Institute of Justice by the Department of Criminology and Criminal Justice.
- Subirats, Joan (1994), *Análisis de políticas públicas y eficacia de la administración*, Madrid, Ministerio para las Administraciones Públicas.

- Valadez, Joseph y Bamberger, Michael (eds.) (1994), *Monitoring and Evaluating Social Programs in Developing Countries: a Handbook for Policymakers, Managers, and Researchers*, Washington, The World Bank (EDI Development Studies).
- Valles, Miguel S. (2007), *Técnicas cualitativas de investigación social: reflexión metodológica y práctica profesional*, Madrid, Editorial Síntesis.
- Wholey, Joseph S.; Hatry, Harry P.; y Newcomer, Kathryn E. (eds.) (2004), *Handbook of Practical Program Evaluation*, San Francisco, Jossey-Bass (Essential Text for Nonprofit and Public Leadership and Management). 2. ed.