

Serie

*Estado, Gestión Pública y
Desarrollo en América Latina*

La evaluación de políticas

Fundamentos conceptuales
y analíticos

Serie

*Estado, Gestión Pública y
Desarrollo en América Latina*

La evaluación de políticas

Fundamentos conceptuales
y analíticos

Título: La evaluación de políticas. Fundamentos conceptuales y analíticos.

ISBN: 978-980-422-053-1

Depósito Legal: DC2016001307

Editor: CAF

Director Corporativo, Desarrollo Institucional
Christian Asinelli

Coordinador de la Serie
Carlos H. Acuña

Compilador del volumen
Roberto Martínez Nogueira

Revisión
Nathalie Gerbasi, Leandro Gorgal

Edición y corrección
Patricia Iacovone

Traducción
Valeria Cardozo

Arte de tapa, diseño y diagramación editorial
Andrés Bermejo Fernández

Las ideas y planteamientos contenidos en la presente edición son de exclusiva responsabilidad de sus autores y no comprometen la posición oficial de CAF.

Esta publicación no puede ser reproducida, archivada o transmitida en ninguna forma o medio alguno, mecánico o electrónico, sin el permiso previo escrito de los titulares del Copyright.

Impreso en Argentina.

© 2016 Corporación Andina de Fomento. Todos los derechos reservados.
www.caf.com

La evaluación de políticas

Fundamentos conceptuales
y analíticos

Autores

Oswaldo Feinstein
Juan Andrés Ligeró Lasa
Martin Rein y Donald Schon
Giandomenico Majone
Carol H. Weiss
Fred Carden y Marvin C. Alkin
Xavier Ballart
Evert Vedung
John Mayne
Robert Picciotto

Índice

| | |
|---|-----|
| Prólogo , por L. Enrique García | 7 |
| Prefacio sobre la Serie , por Christian Asinelli y Carlos H. Acuña | 11 |
| Introducción. Fundamentos conceptuales y analíticos de la evaluación de políticas | 15 |
| El sendero de complejización de la evaluación | |
| <i>La evaluación pragmática de políticas públicas</i> Osvaldo Feinstein | 27 |
| El sendero epistemológico y el encuentro de la teoría con la realidad | |
| <i>Dos métodos de evaluación: criterios y teoría del programa</i> Juan Andrés Ligeró Lasa | 49 |
| <i>Un discurso de políticas que refleja su marco</i> Martin Rein y Donald Schon | 111 |
| <i>Programas de investigación y programas de acción o ¿puede la investigación de políticas aprender de la ciencia?</i> Giandomenico Majone | 147 |
| <i>La investigación de políticas: ¿datos, ideas y argumentos?</i> Carol H. Weiss | 169 |
| <i>Raíces de la evaluación: una perspectiva internacional</i> Fred Carden y Marvin C. Alkin | 205 |
| La diversidad teórica y los modelos de práctica | |
| <i>Modelos teóricos para la práctica de evaluación de programas</i> Xavier Ballart | 229 |
| <i>Investigación de la evaluación</i> Evert Vedung | 267 |
| <i>Abordaje de la atribución a través del análisis de contribución: uso sensato de las mediciones de desempeño</i> John Mayne | 305 |
| <i>La independencia de la evaluación en las organizaciones</i> Robert Picciotto | 331 |

Prólogo

*L. Enrique García, presidente ejecutivo
CAF -Banco de Desarrollo de América Latina*

Afirmar que las instituciones presentan una importancia superlativa en el desarrollo de nuestros países constituye una premisa que actualmente exhibe un cierto nivel de consenso, tanto en los círculos de investigación académica como en las comunidades de práctica de política pública. En cambio, definir *qué* son las instituciones y *cuáles* son los aspectos que nos importan de ellas abre un abanico de enfoques y explicaciones de gran diversidad en el campo de las ciencias sociales.

En esa línea, la comprensión de los problemas del desarrollo en nuestra región enfrenta tres obstáculos que, a mi entender, ostentan una relevancia estratégica. Primero, el debate y la formación sobre los determinantes del desarrollo tienden, en general y por diversas razones, a perder de vista un largo proceso de producción de conocimiento sobre el papel que le cabe a los factores político-institucionales, así como a la forma en que los conceptualizamos. En consecuencia, la búsqueda de un conocimiento y una capacitación *de punta* sobre el desarrollo latinoamericano corre el riesgo de llevarse adelante sin la suficiente contextualización histórica o sin el aprovechamiento de un valioso —y necesario— trabajo de elaboración conceptual.

Segundo, y de la mano del primero, muchas veces este debate se ve *apropiado* por una disciplina o enfoque que, en la búsqueda de simplificación y de parsimonia explicativa, usualmente desestima la

presencia de variables que, a la larga, los procesos históricos nos imponen como ineludibles al momento de comprender la lógica de los desafíos y de diseñar e implementar las estrategias necesarias para el alcance y la sostenibilidad del desarrollo en la región.

Tercero, y más allá de la mayor circulación de información que existe en la actualidad, las barreras idiomáticas y regionales todavía dificultan el aprovechamiento, por parte del debate y la formación en América Latina, de las novedades o los análisis relevantes para la comprensión de la economía política de los procesos de desarrollo.

Con el ánimo de contribuir a la superación de estos obstáculos y en el marco de una estrategia general de expansión de la oferta de distintas modalidades de capacitación orientadas a las gerencias públicas de los países de nuestra región, CAF impulsa la publicación de esta serie de volúmenes sobre *Estado, Gestión Pública y Desarrollo en América Latina*. Su objetivo general es promover la creación de una cultura generalizada de excelencia en la gestión pública y compartir experiencias que permitan apoyar a quienes participan activamente en el proceso de diseño y ejecución de políticas públicas asociadas al fortalecimiento institucional.

En términos específicos, esta Serie persigue:

- a) Rescatar el conocimiento que se encuentra detrás de los vaivenes teóricos relacionados con la conceptualización y la formación ligados a la naturaleza y papel del Estado, las instituciones, la política y las políticas públicas.
- b) Brindar elementos para un debate que facilite el establecimiento de miradas interdisciplinarias, con capacidad de encarar, de forma más robusta, las complejidades de los procesos de desarrollo.
- c) Ofrecer a los especialistas en estos campos temáticos, de manera unificada para su análisis y sus tareas de formación, textos que, aunque conocidos, se encuentran usualmente dispersos y en diversos idiomas, con la finalidad de presentar en español algunas de estas valiosas novedades que se han publicado en las últimas décadas.

Por último, quiero agradecer la disposición de los autores, así como de las revistas y editoriales en los que los trabajos aquí incluidos fueron publicados originalmente. También quiero destacar el compromiso de la Dirección Corporativa, Desarrollo Institucional de CAF, así como de los expertos que han colaborado en los procesos de selección y edición de estos volúmenes.

Sin duda, esta exhaustiva serie de textos fundamentales en materia de desarrollo y políticas públicas constituirá un gran aporte para enriquecer la gestión del conocimiento en esta temática y así contribuir al debate y a una mayor comprensión de los principales retos que la institucionalidad regional tiene para el futuro.

Prefacio sobre la Serie

Estado, Gestión Pública y Desarrollo en América Latina

Christian Asinelli y Carlos H. Acuña

El prólogo de Enrique García nos introduce a los obstáculos que enfrentamos para la comprensión de los problemas del desarrollo y a la consecuente necesidad de (a) rescatar de posibles olvidos el conocimiento que se encuentra detrás de los vaivenes teóricos relacionados con la conceptualización y la formación ligados a la naturaleza y papel de Estado, las instituciones, la política y las políticas públicas; (b) brindar elementos para un debate que facilite el establecimiento de miradas interdisciplinarias, con capacidad de encarar, de manera más robusta, las complejidades de los procesos de desarrollo; (c) para el público constituido por los especialistas en los campos temáticos cubiertos por la serie, ofrecer textos de manera unificada para su análisis y tareas de formación que, aunque conocidos, se encuentran usualmente dispersos y en diversos idiomas, facilitando en español algunas de estas valiosas novedades que se han publicado en otros idiomas en las últimas décadas.

En este contexto, deseamos profundizar algunos aspectos que le han dado cuerpo a la serie *Estado, Gestión Pública y Desarrollo en América Latina* desde la Dirección Corporativa, Desarrollo Institucional de CAF. Las políticas públicas que sostienen el desarrollo en América Latina han experimentado profundos cambios y redefiniciones del papel y la interrelación entre sociedad, Estado y mercado, así como de su inserción en el contexto internacional en las últimas décadas. Por ello, a los obstáculos al desarrollo enfrentados históricamente se ha sumado el reto de una

realidad —nacional, regional y global— cuya veloz mutación presenta en simultáneo oportunidades y problemas novedosos, lo que implica que su significado para el desarrollo latinoamericano resulta muchas veces incierto. Este proceso de rápidos cambios y consecuente incertidumbre mina la sostenibilidad de miradas y acciones con horizontes de largo plazo, por lo que, en definitiva, desafía nuestro entendimiento y el diseño e implementación de estrategias de desarrollo.

La mayor o menor capacidad de las políticas públicas de desarrollo para alcanzar el difícil equilibrio entre cambios y continuidades —equilibrio ineludible en la respuesta a necesidades presentes sin perder de vista el largo plazo— depende de las características, debilidades y fortalezas de las dinámicas político-institucionales que producen esas políticas. Hace tiempo que CAF reconoció el papel estratégico que le cabe a estos procesos político-institucionales y a la capacitación que permite una mejor comprensión y el fortalecimiento de estos. Por ello, esta serie se inscribe como una pieza más en el esfuerzo que, desde hace ya quince años, CAF sostiene en alianza con contrapartes académicas en siete países latinoamericanos y con más de 52.000 participantes: los Programas de Gobernabilidad, Gerencia Política y Gestión Pública y de Liderazgo para la Transformación.

Como parte de esta dinámica, esta serie apunta a difundir y cimentar la comprensión e implicancias que el funcionamiento del Estado y de las instituciones en América Latina tienen para el desarrollo de nuestra región, brindando información sobre los fundamentos del debate teórico internacional, así como también transmitiendo el análisis de experiencias significativas de nuestras sociedades y de otras latitudes.

CAF, como institución dedicada a promover el desarrollo, pone esta colección a disposición no solo de sus Programas de formación, sino también de todos los que participan en la gestión pública, así como de académicos, estudiantes y los interesados y comprometidos con los problemas y soluciones de nuestra región, con el propósito de acercarles esta literatura y de apoyarlos en el proceso de comprensión, diseño y ejecución de políticas públicas asociadas a la calidad institucional y al desarrollo en nuestros países.

Para concluir este breve prefacio, nos resta agradecer a todos los que han hecho posible esta serie. En primer lugar, al presidente ejecutivo de CAF, Enrique García, por su amplitud y visión estratégica al promover la realización de estos volúmenes y su interés en impulsar la formación del capital humano, consciente de que los resultados de los procesos de capacitación son lentos, pero con la convicción de que este esfuerzo redunda en una mejora de la calidad institucional de la región.

Queremos sumar nuestro reconocimiento a las diferentes áreas de CAF que llevaron delante de manera exitosa esta serie: la Dirección de Administración de Fondos de Cooperación para el Desarrollo (DAFCD), la Consultoría Jurídica, la Dirección de Comunicación Estratégica (DCE) y el Centro de Información y Documentación (CID). También, al equipo de la Dirección Corporativa, Desarrollo Institucional, Nathalie Gerbasi y Leandro Gorgal, por su dedicación y empeño.

Finalmente, deseamos expresar nuestra gratitud a los especialistas externos que colaboraron en las investigaciones y compilaciones para cada uno de los volúmenes, así como a las instituciones, editoriales y autores que contribuyeron a la multiplicación del conocimiento plasmado en estos libros.

Introducción. Fundamentos conceptuales y analíticos de la evaluación de políticas[†]

Este volumen pretende poner a disposición del lector temas y enfoques referidos a la evaluación de políticas y programas públicos. Su propósito es presentar y discutir fundamentos epistemológicos, conceptuales y analíticos dejando, en esta instancia, en un segundo plano las dimensiones metodológicas y operacionales. La atención está centrada en los desafíos intelectuales que la evaluación plantea y en las principales cuestiones que fueron emergiendo y consolidándose desde los primeros intentos de someter a examen sistemático la formulación, la implementación y las consecuencias efectivas de las políticas de gobierno, hasta llegar a convertirse en el escenario de disputas de gran relevancia para el desarrollo del conocimiento sobre la acción social y sobre sus estrategias y efectividad.

Esta problemática puede abordarse considerando distintos ejes de reflexión que se entrecruzan y condicionan mutuamente. El volumen trata solo algunos de ellos pues no tiene ambición de exhaustividad. Si este fuese el objetivo, la enorme variedad de enfoques, teorías y criterios operativos haría vano el esfuerzo. El criterio de selección consistió en incluir textos que son o hacen referencia a contribuciones fundamentales para el desarrollo y la comprensión de cuestiones reiteradamente abordadas en este campo, varios de los cuales no son fácilmente accesibles al público de habla española.

† Roberto Martínez Nogueira es el autor de esta introducción.

Esta introducción presenta algunas consideraciones generales y explicita el sentido de la selección de los textos. La evaluación de políticas públicas es a la vez una promesa, una exigencia y un campo en el que se despliegan concepciones y visiones diversas sobre el conocimiento y su articulación con la práctica social. Su comprensión como instrumento de decisión y de gestión dio, desde su origen, lugar a múltiples debates en torno a su sentido y utilidad, así como a la pertinencia de apelar a distintas tradiciones intelectuales para el desarrollo de marcos analíticos y de orientaciones metodológicas de una temática que se ha convertido en una disciplina con elevada autonomía y que ha dado lugar a una orientación profesional consolidada. Su institucionalización es creciente, tal como se manifiesta en su desarrollo académico, en la conformación de comunidades profesionales y en su incorporación a sistemas de planificación, presupuestación e innovación de la gestión.

Como promesa, la evaluación es visualizada como un medio potente para enriquecer el debate público y para promover y dar rigurosidad al aprendizaje colectivo sobre iniciativas, estrategias y acciones dirigidas a incidir sobre las políticas y la organización y el funcionamiento estatal, que brinda elementos de juicio para la definición de prioridades, para la elección de medios alternativos para alcanzar objetivos sociales y para alimentar la participación y el control social. Su carácter de exigencia deriva de sus aportes asociados a la transparencia de la acción del Estado, a la calidad de la acción de gobierno y a la postulación democrática de una ciudadanía alerta, informada y activa. Como campo, la evaluación está integrada por prácticas diversas, construida por aportes de cuerpos de conocimientos múltiples y experiencias muy distintas en procesos polémicos y disputados, sin consensos medianamente establecidos sobre fundamentos epistemológicos, metodológicos y operativos relativos a la articulación entre conocimiento, poder y acción.

Naturaleza y desafíos de la evaluación

La evaluación plantea desafíos muy específicos. Algunos son propios de la naturaleza política de los múltiples juegos en que intervienen los actores sociales en torno a las políticas públicas; otros corresponden a la problemática del conocimiento que es movilizado a lo largo de los procesos

de definición de problemas, de generación y consideración de alternativas, de formalización y de gestión de las políticas y de identificación y apreciación de sus consecuencias.

En primer lugar, las políticas son construcciones sociales producto de procesos orientados a la superación de situaciones públicamente relevantes frente a las que diversos agentes despliegan estrategias respaldadas por recursos de poder en arenas de colaboración y negociación, con permanentes tensiones y conflictos. Son realidades dinámicas cuyos contenidos, ejecución, resultados e impactos dependen de condiciones institucionales, de la incidencia del contexto y de las capacidades políticas, técnicas y de gestión de los actores involucrados.

Además, las políticas hacen uso de conocimientos parciales o limitados, expresándose a través de instrumentos cuya efectividad es un interrogante. La formulación de las políticas se apoya en relaciones de causalidad hipotéticas, con un entendimiento problemático y limitado de las situaciones a superar. De igual manera, las estrategias de intervención están fundadas en medios, conocimientos y comprensiones parciales, con probabilidades elevadas de que los objetivos enunciados solo resulten guías imperfectas para orientar la ejecución.

Aun cuando un buen diseño y una adecuada implementación incrementan la probabilidad de que resultados e impactos se acerquen a los objetivos buscados, de ninguna manera ofrecen certeza sobre su logro. Asimismo, las políticas no pueden ser analizadas o evaluadas exclusivamente por los enunciados de sus intencionalidades, las retóricas que las legitiman o las leyes, reglamentos, decretos o disposiciones de tipo administrativo que les dan vigencia: estos son partes constitutivas de las políticas, pero su sentido efectivo se construye a través de decisiones, regulaciones, actividades, recursos, contenidos simbólicos y comportamientos que conforman las acciones de interpretación, aplicación e implementación dirigidas a alcanzar resultados o impactos en la sociedad.

En este escenario de incertidumbres generalizadas y de poder limitado, la evaluación debe identificar las brechas existentes entre las intenciones y las consecuencias, reconociendo los factores que inciden sobre los resultados

o determinan éxitos o fracasos, ofreciendo explicaciones sustentadas en evidencias que permitan perfeccionar las argumentaciones y alimentar el aprendizaje colectivo. Es, también, un medio para el fortalecimiento de capacidades políticas y de gestión cuyo resultado se manifiesta en la superación de las debilidades o deficiencias de los diseños de políticas y acciones, en una mejor implementación y en estrategias renovadas para superar los retos contextuales.

Un segundo desafío es consecuencia de la necesidad de proceder a recortar la realidad, delimitando el ámbito considerado. Las políticas no impactan la realidad aisladamente, sino que conforman conjuntos de gran diversidad de políticas que se expresan en intervenciones o acciones con propósitos muy específicos, dirigidas a destinatarios o receptores varios —ciudadanos, poblaciones, actividades o territorios particulares— y que afectan, de manera diferencial y localizada, sus esferas de acción, condiciones de vida, recursos y oportunidades. Desde ya, por razones cognitivas y operacionales, la evaluación no puede referirse al conjunto de las políticas, ni abarcar a la totalidad de los efectos de la acción pública, ni atender a todas las dependencias e interacciones mutuas. Esto agrega complejidad al análisis, generando uno de los retos más importantes de la tarea de evaluación: la adecuada atribución, es decir, la adjudicación de un resultado o consecuencia a la política y a los elementos que componen su estrategia de intervención, discriminándolos de aquellos que pueden estar operando en el contexto de la política, pero que no constituyen parte de ella o del programa (aspecto que merece un capítulo especialmente dedicado al tema).

En tercer lugar, lo anterior pone de manifiesto otro desafío propio de la práctica misma. La disciplina de la evaluación no ha generado ni un lenguaje ni un cuerpo teórico de aceptación unánime, ya que se ha consolidado recurriendo a diferentes concepciones sobre las decisiones y los procesos políticos. Si bien hay acuerdo sobre su exigencia básica de aplicación sistemática de los patrones y procedimientos de la investigación social, este enunciado no da cuenta de las opciones cognitivas, normativas e instrumentales involucradas. La multitud de estas opciones ha dado lugar a la proliferación de enfoques y a un pluralismo de estrategias, diseño, metodologías y de planteo de la ejecución. Ello ha provocado

que las distintas opciones confieran diferente sentido e interpretación a la conversión de los resultados de los análisis en recomendaciones políticas y técnicas sobre la eficacia, la eficiencia, las consecuencias distributivas o el impacto de las políticas evaluadas.

Cuarto, las políticas integran redes con relaciones complejas de ordenamiento y complementariedad que dan lugar a cuestiones muy relevantes referidas a la coherencia de orientaciones y contenidos, y de coordinación a lo largo de sus ciclos de vida. Este es un aspecto poco desarrollado en la literatura sobre evaluación, pues la atención se ha focalizado en políticas singulares con escasa consideración de sus interdependencias con otras. Este desafío está ganando en visibilidad e importancia, pues existe una brecha creciente entre la naturaleza de los problemas que demandan la atención pública —pobreza, empleo, innovación, hábitat, desarrollo territorial, etc.— y las acciones dirigidas a superarlos, tanto por su carácter multidimensional, por los límites difusos entre distintos ámbitos de lo social o por las insuficiencias manifiestas de los mecanismos institucionales para encarar de manera integral estas cuestiones. Obviamente, este desafío remite a los marcos institucionales y a los mecanismos organizaciones para el diseño y la implementación de políticas.

El desarrollo de la evaluación

La evaluación es mucho más que la aplicación de conocimiento científico para determinar las consecuencias de los cursos de acción adoptados. Esta aplicación confronta límites nacidos de la imperfecta disponibilidad de información y de la frecuente falta de coherencia entre la naturaleza de los procesos y comportamientos estudiados y las metodologías utilizadas, y condiciona la validez de las conclusiones. El desarrollo de la evaluación se explica por la utilización de diferentes estrategias y técnicas para superar estos problemas, con una tensión permanente entre las aproximaciones cuantitativas y cualitativas que suponen distintas concepciones del conocimiento y de la acción social.

En el desarrollo histórico de esta práctica, pueden identificarse hitos particulares y secuencias que son reflejo de cambios en los procesos políticos, de nuevos papeles asumidos por las ciencias sociales, de nuevos

modos de pensar las decisiones públicas y la gestión estatal y de contextos de mayor rigurosidad fiscal. Este desarrollo implica esfuerzos de distinta naturaleza por dar mayor efectividad y eficiencia a la acción estatal a través de la incorporación de instrumentos analíticos durante los procesos de decisión en torno a las políticas públicas. Esos esfuerzos se dieron primero de forma puntual, no sistemática ni generalizada, utilizando como paradigma básico el método experimental. Puede señalarse como uno de los antecedentes más interesante en el desarrollo de la evaluación el debate en torno a ciertos programas innovadores dirigidos a la problemática de la pobreza durante la década de los sesenta. Este debate fue resuelto con el compromiso asumido por el Poder Ejecutivo ante el Congreso de los Estados Unidos de realizar evaluaciones para determinar la efectividad de acciones sin antecedentes y con sustento teórico y operativo insuficiente. A partir de este episodio, las evaluaciones adquirieron el carácter de prácticas usuales, tanto para fundamentar el diseño de programas como para cuestionar sus resultados y consecuencias. La creciente institucionalización se expresó en la formación de organizaciones académicas y profesionales, en la proliferación de congresos y revistas especializadas y en la generalización del uso de la evaluación como componente del llamado «buen gobierno», alentado todo ello por una transferencia de políticas y de modelos de gestión realizada activamente por la comunidad internacional de agencias de desarrollo para la instalación de la evaluación en los poderes del Estado y en las organizaciones sociales concernidas con lo público.

A lo largo de este proceso, muchos de los postulados derivados del paradigma experimental fueron revisados, adecuándolos a la creciente complejidad de las políticas debido a la multiplicidad y diversidad de actores que se movilizan en torno a ellas, al pluralismo de orientaciones estratégicas y a la incertidumbre propia de los comportamientos de los actores y de los procesos sociales. Desde acciones puntuales que obedecían a demandas específicas de los responsables de las políticas o de los mecanismos de financiamiento, se fue avanzando hacia la incorporación regular y sistemática de la evaluación en los diseños de políticas, programas y proyectos, con marcos normativos y organizaciones con competencias para su promoción, regulación y construcción de capacidades. De igual manera, fueron ganando legitimidad enfoques competitivos más pragmáticos y con aspiraciones de mayor realismo, atendiendo tanto a las condiciones

efectivas de producción, disponibilidad y uso de la información, a las particularidades del contexto de la acción y a la conveniencia de disponer de una batería más compleja y diversa de criterios de evaluación.

Los nuevos modelos de gestión pública pusieron énfasis no tanto en los procesos programables y regulares de producción administrativa, sino en sus resultados e impactos. La gestión por resultados, de amplia expansión desde hace unas pocas décadas, conlleva lógicamente la necesidad de identificar el grado en que los objetivos de gobierno son alcanzados, con consideraciones de eficiencia y distributivas. Hoy, la evaluación forma parte de los requerimientos de los ministerios encargados de los presupuestos, de las condicionalidades de los organismos internacionales de financiamiento, de las «buenas prácticas» transferidas por medio de organizaciones académicas y profesionales, siendo las metodologías y técnicas difundidas de las maneras más diversas. Constituye, junto con la planificación y el monitoreo, un instrumento esencial para que la gestión política y de gobierno respondan a principios de coherencia, evidencia, transparencia y acumulación de aprendizaje.

El sendero de complejización de la evaluación

La colección de estudios es introducida por una contribución de interés por su sistematización de conceptos, de los modos operativos y del proceso de institucionalización de la evaluación. El trabajo de Osvaldo Feinstein, «La evaluación pragmática de políticas públicas», anticipa cuestiones que son profundizadas en otros capítulos de este volumen. Este trabajo tiene un valor especial, pues apunta a antecedentes conceptuales y analíticos inevitables para comprender el sentido y la evolución de esta temática.

El sendero epistemológico y el encuentro de la teoría con la realidad

El trabajo de José Andrés Ligeró Lasa, «Dos métodos de evaluación: criterios y teoría del programa», presenta dos enfoques relativamente recientes. Estos someten a examen la teoría de la acción sobre la que se construye la política o el programa, y procuran una explicitación mayor de las consecuencias de los diseños de las políticas y de su articulación lógica con la implementación y la evaluación.

Luego de esta presentación, se incluye un conjunto de trabajos que plantean interrogantes epistemológicos y teóricos. Martin Rein y Donald Schon, en «El discurso que refleja su marco», discuten la relevancia de los marcos analíticos que fundamentan las controversias en torno a las políticas y brindan elementos para entender los sentidos y alcances de los conflictos entre perspectivas.

Como continuación de esta temática, se presenta un análisis de la relación entre conocimiento y políticas elaborado por Giandomenico Majone, «Programas de investigación y programas de acción o ¿puede la investigación de políticas aprender de la ciencia?». La perspectiva adoptada es la de la filosofía de la ciencia e introduce la polémica sobre si la evaluación como tipo específico de investigación difiere o no de la investigación científica convencional, como sostienen los posicionamientos que remiten al positivismo. Este trabajo anticipa una línea de confrontación entre perspectivas que se inició hace décadas y que en el presente adquiere mayor visibilidad y relevancia por el giro paradigmático que se ha producido en las ciencias sociales que, por supuesto, alcanza a la evaluación, y da lugar a un tránsito de las posturas racionalistas a otras de carácter interpretativo.

Por su parte, la contribución de Carol Weiss, «La investigación de políticas: ¿datos, ideas y argumentos?», es sumamente ilustrativa sobre el papel y relevancia normativa tanto de los insumos como de los productos de la actividad de análisis y evaluación, resultante y generadora a la vez de datos, ideas y argumentos en los ciclos de políticas.

La quinta contribución en esta línea «Raíces de la evaluación: una perspectiva internacional», de Fred Carden y Marvin Alkin, tiene una inclusión intencionada. Por una parte, estos autores, importantes por sus aportes a la construcción de teorías, pasan revista y categorizan los marcos teóricos disponibles. Por otra, presentan contribuciones «periféricas» en países de ingresos medios que se apartan de los cánones establecidos.

La diversidad teórica y los modelos de práctica

La sucesión de modelos teóricos con gran influencia sobre la práctica de la evaluación es el contenido del trabajo de Xavier Ballart «Modelos teóricos para la práctica de evaluación de programas». Se identifican los aportes de referentes que contribuyeron a construir y dar fundamento a esta «disciplina científica», conformando modelos teóricos con pretensiones de servir de marco para la tarea de evaluación.

El trabajo de Evert Vedung, «Investigación de la evaluación», constituye una presentación sistemática del estado actual del desarrollo de esta práctica, como es convencionalmente incorporada a los programas de formación, con fuerte adhesión de la comunidad profesional y adoptada como norma por muchas organizaciones con funciones de evaluación de políticas públicas.

El trabajo de John Mayne, de la Auditoría General de Canadá, «Abordaje de la atribución a través del análisis de contribución: uso sensato de las mediciones de desempeño», expone un método para abordar la problemática de la atribución, cuestión de gran importancia para fundamentar conclusiones y formular recomendaciones dirigidas al perfeccionamiento de las políticas y programas. Su abordaje es, a la vez, claro y potente conceptualmente, así como útil para llevar a cabo análisis contribuyentes a una mayor validez y rigurosidad. Las referencias a las mediciones asociadas a la gestión por resultados son ilustrativas de las convergencias necesarias entre gestión y evaluación.

El artículo de Robert Picciotto, «La independencia de la evaluación en las organizaciones», figura importante por su influencia en los debates sobre la contribución de la evaluación al perfeccionamiento de las intervenciones dirigidas al desarrollo, presenta el conflicto muchas veces enfrentado por los evaluadores que actúan en contextos organizacionales. Este es un conflicto entre la ética profesional y las demandas de la jerarquía que puede llevar a elecciones inadecuadas de enfoques, metodologías y técnicas de evaluación, con conclusiones con problemas de validez.

A modo de síntesis

Los trabajos incorporados a esta selección de lecturas brindan, conforme al propósito enunciado de ilustrar la complejidad del campo y de la práctica de la evaluación, un panorama no exhaustivo pero suficientemente amplio. Plantean, además, los desafíos en adelante en materia de avance teórico y de mayor relevancia y utilidad de sus aportes al aprendizaje social y al perfeccionamiento de las políticas públicas.

Por cierto, el volumen constituye un punto de partida que espera contribuir a enriquecer las intervenciones dirigidas a la institucionalización de la evaluación a lo largo de un sendero que necesariamente deberá ser recorrido como consecuencia de los imperativos políticos y éticos de mejorar la eficacia del Estado y de satisfacer las demanda de democratización de las políticas.

La evaluación se ha convertido en un campo de gran vitalidad, con intensos debates disciplinarios, teóricos y metodológicos que lo han enriquecido. No obstante estos avances, la comprensión de los desafíos que esta práctica enfrenta no se manifiestan de manera semejante en los diferentes países, ámbitos sectoriales y organizaciones de gobierno. Puede afirmarse que el escenario actual en América Latina aún requiere la consolidación de su institucionalización y una mayor utilización de sus resultados en el diseño y la implementación de políticas y programas públicos.

No se desconoce que esa institucionalización es un proceso históricamente situado que responde a las especificidades de cada contexto de acción, pero, a la vez, este volumen constituye un mensaje que reafirma que la evaluación tiene un carácter muy exigente en materia de definición de los problemas y de los objetivos que las políticas pretenden, respectivamente, superar o alcanzar, así como en cuanto a la calidad de su diseño, a la confiabilidad de los datos que se utilizan y a la rigurosidad y a la aspiración de validez en la fundamentación de las conclusiones, con viabilidad y consistencia lógica de las recomendaciones que se formulan.

La adecuación a circunstancias particulares no debe llevar a distorsionar la aspiración de que la evaluación brinde conclusiones relevantes fundadas en evidencias sistemáticamente recogidas y rigurosamente analizadas, ni a servir como excusa para convertirla en argucia retórica legitimadora de acciones estatales. La evaluación debe permitir, como lo argumentan algunos de los trabajos incluidos en este volumen, una deliberación informada para recuperar y permitir la acumulación de los aprendizajes sociales en torno a las políticas públicas.

Notas

¹ La selección de artículos para este volumen fue llevada a cabo por Roberto Martínez Nogueira para la presente serie Estado, Gestión Pública y Desarrollo en América Latina, dirigida por Christian Asinelli y coordinada por Carlos H. Acuña.

Referencias

Ballart, Xavier, «Modelos teóricos para la práctica de la evaluación de programas», en Quim Brugue y Joan Subirats (comp.) *Lecturas de gestión pública*, Ministerio de las Administraciones Públicas de España, Madrid, 1999.

Carden Fred y Alkin Marvin C., «Evaluation Roots: An International Perspective», en *Journal of MultiDisciplinary Evaluation*, Volumen 8, N.º 17, Enero 2012.

Feinstein, Osvaldo, *La evaluación pragmática de políticas públicas*, Evaluación de políticas públicas, Revista ICE, Madrid, Mayo-Junio 2007, N.º 836, pp. 9-31.

Ligero Lasa, Juan Andrés, *Dos métodos de evaluación: criterios y teoría del programa*, Madrid, CEU Ediciones del Centro de Estudios de la Cooperación, Fundación Universitaria San Pablo, 2011.

Majone, Giandomenico, «Programas de investigación y programas de acción o ¿puede la investigación de políticas aprender de la ciencia?», en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.) *Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas*, Fondo de Cultura Económica, México, 1999, pp. 358-376.

Mayne, John, «Addressing attribution through contribution analysis: using performance measures sensibly», *The Canadian Journal of Program Evaluation*, Ottawa, Canadian Evaluation Society, Vol. 16, 1, 2001, pp. 1-24.

Picciotto, Robert, «Evaluation independence in organizations», *Journal of MultiDisciplinary Evaluation*, Volumen 9, N.º 20, 2013.

Rein, Martin y Schon, Donald, «El discurso que refleja su marco», en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.) *Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas*, Fondo de Cultura Económica, México, 1999, pp. 327-356.

Vedung, Evert, «Evaluation Research», en Guy Peters y John Pierre (comp.), *Handbook of public policy*, Sage Publications, Londres, 2006, pp. 397-416.

Weiss, Carol, «La investigación de políticas: ¿datos, ideas y argumentos?», en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.) *Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas*, Fondo de Cultura Económica, México, 1999, pp. 377-406.

La evaluación pragmática de políticas públicas*

Osvaldo Feinstein**

1. Introducción

El enfoque tradicional de la evaluación ha estado centrado en proyectos y en la fase de decisión relacionada con su aprobación, es decir, con la evaluación *ex ante* de proyectos. Este enfoque se ha ido generalizando y adaptando para aplicarlo a programas y a políticas públicas que han pasado por la fase de implementación. El énfasis en una gestión orientada a resultados y la necesidad de rendir cuentas por el uso de los recursos empleados, así como de aprender de la experiencia, ha generado un creciente interés en la evaluación como una disciplina que contribuya a estas demandas.

2. Funciones y tipos de evaluaciones

La evaluación se lleva a cabo con dos propósitos principales: para aprender de la experiencia, extrayendo lecciones que puedan aplicarse con el fin de mejorar el diseño y la implementación de políticas públicas (y/o programas y/o proyectos), y para la rendición de cuentas al público y/o a quienes han aportado los recursos utilizados, y en algunos casos al Parlamento. El primer propósito se cumple con las llamadas evaluaciones formativas, cuya finalidad es el aprendizaje, en tanto que el segundo propósito caracteriza a las llamadas evaluaciones sumativas, que formulan un juicio sobre el valor de lo evaluado.

Por otra parte, y según el momento en que se realizan las evaluaciones, se puede distinguir entre evaluación *ex ante*, que corresponde al proceso de diseño de las políticas públicas (u otras intervenciones) y que consiste en una consideración de la relevancia o pertinencia de la intervención y de una estimación de sus posibles efectos. Cabe destacar que no siempre se hacen este tipo de evaluaciones, sobre todo cuando hay prisa por aprobar políticas como señales o indicaciones de que se está haciendo algo en relación con un problema que la población percibe como importante (por ejemplo, una política de inmigración). Cuando hay una fecha de terminación de una política pública (por ejemplo, una *sunset clause*), las evaluaciones que se realizan después de esta fecha se denominan evaluaciones terminales o *ex post*¹, en tanto que las que tienen lugar después del inicio de la implementación, pero antes de su conclusión son evaluaciones intermedias, que suelen utilizarse para aportar elementos de juicio a la hora de decidir sobre la continuación de una política. Más adelante se presentará otra distinción entre evaluaciones de procesos y evaluaciones de impacto.

3. Evaluación y evidencia

Desde fines de los años noventa se ha ido enfatizando el diseño de políticas públicas con base en evidencias (*evidence-based policies*), lo cual se ha percibido como un impulso renovado a la evaluación como fuente de evidencia. Sin embargo, cabe destacar que ni toda la evidencia surge de evaluaciones (hay investigaciones que no son evaluaciones y que constituyen una importante fuente de generación de evidencias), ni toda evaluación genera evidencias (a veces las evaluaciones son una descripción de la implementación y carecen de evidencia sólida).

Por otro lado, la evidencia no es la única base sobre la cual se diseñan políticas públicas. El proceso de formulación de políticas públicas es altamente complejo y si bien las evidencias sobre resultados son un insumo importante, especialmente en algunas circunstancias, existen factores políticos que también tienen una fuerte influencia en el proceso de diseño².

4. ¿Cómo permite la evaluación aprender de la experiencia?

La evaluación permite aprender de la experiencia extrayendo lecciones

sobre lo que ha funcionado bien y lo que ha fallado, y las razones de éxitos y fracasos, propios y ajenos.

Si las políticas públicas se implementaran siempre de acuerdo a las intenciones, y si no hubiera efectos no buscados, la evaluación no sería útil, excepto para confirmar el logro de los objetivos. En ese tipo de mundo no habría incertidumbre, coincidiendo lo planeado con lo realizado (no habría diferencias entre *ex ante* y *ex post*). Pero en el mundo real las políticas no siempre pueden implementarse conforme a las intenciones, y los planes no siempre se cumplen, o se cumplen solo parcialmente. Además, frecuentemente, se presentan efectos no buscados, positivos y/o negativos. Por lo tanto, es importante verificar qué es lo que realmente ha sucedido y el grado en que se han cumplido los objetivos, lo cual permitirá mejorar la eficacia y eficiencia de las intervenciones futuras, posibilitando un mejor uso de los recursos para alcanzar un mayor nivel de resultados positivos y evitar o mitigar los efectos indeseados negativos.

Hay diferentes tipos de evaluaciones que pueden utilizarse para aprender de la experiencia. Entre ellas cabe destacar las evaluaciones de procesos (que focalizan la atención en los mecanismos e instrumentos a través de los cuales se ha intentado alcanzar determinados objetivos) y las evaluaciones de impacto (que se centran en los resultados últimos de las políticas públicas, esto es, en las modificaciones logradas en un conjunto de variables-objetivo, por ejemplo, en los componentes del desarrollo humano o en los objetivos del programa de gobierno).

Si bien lo que realmente interesa es el impacto de las políticas o programas o proyectos, y, por lo tanto, las evaluaciones de impacto, hay circunstancias en las cuales no es posible determinar de modo inequívoco cuál ha sido el efecto atribuible a determinadas intervenciones y, por ello, es importante complementar el análisis de impacto con un análisis de procesos.

Volveremos sobre estas cuestiones más adelante. Lo que vale la pena tener presente es que la evaluación puede y debe ayudar a evitar el riesgo que más acecha a quienes desean aprender de la experiencia: la falacia *post hoc ergo propter hoc* (que puede traducirse casi literalmente como «después de algo, por lo tanto debido a eso»), confundiendo las causas con lo que

ha sucedido primero. Por ejemplo, si se implementó un proyecto de desarrollo rural en una región de un país y aumentaron los ingresos de la población rural en dicha región, esto no implica necesariamente que los ingresos aumentaron debido al proyecto ya que, por ejemplo, el efecto positivo pudo haberse debido a un fuerte incremento de los precios de los productos producidos en esa región por razones independientes al proyecto (debido a una devaluación o a una fuerte inmigración o a un fenómeno climático que destruya la producción en otras regiones). La evaluación debe tomar en cuenta los factores contextuales y realizar un análisis y síntesis que permita identificar las causas que han operado para alcanzar o no los resultados esperados, así como otros posibles efectos atribuibles a las intervenciones que tuvieron lugar. De este modo será posible extraer lecciones de la experiencia que puedan, en principio, aplicarse en contextos similares. El desafío, entonces, es cómo diseñar, implementar y utilizar evaluaciones de modo que permitan aprender de la experiencia, tema que será abordado en el resto de este trabajo (y en las referencias que se incluyen al final de este).

5. Criterios de evaluación

La evaluación lleva a considerar sistemáticamente varios factores, los cuales a veces se presentan como principios o criterios de la evaluación: la relevancia o pertinencia de lo que se evalúa, la eficacia o efectividad (el grado de logro de los objetivos) y la eficiencia (en el uso de los medios para alcanzar los objetivos).

A veces, se incorporan también como criterios explícitos la sostenibilidad de los resultados, el desarrollo institucional y la coherencia de las políticas o programas o proyectos³. Sin embargo, la sostenibilidad puede incluirse en la eficacia o efectividad (ya que si los resultados no son sostenibles, la eficacia sería solo de corto plazo), en tanto que el desarrollo institucional o es un objetivo o es un efecto no buscado; finalmente, la coherencia puede subsumirse en la relevancia o pertinencia de la intervención (política, programa o proyecto). Aplicando de esta forma la «navaja de Ockham», los criterios evaluativos se reducirían a los tres mencionados en el párrafo precedente⁴. No obstante, dependiendo del contexto político, en algunos casos puede tener sentido incorporar criterios adicionales como la equidad,

cuya introducción permitiría en la evaluación tomar en cuenta cuestiones que sean consideradas relevantes⁵.

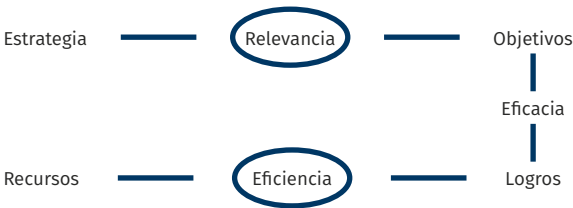
La relevancia de una intervención se establece relacionando sus objetivos con la estrategia del país, sector, organización o institución que haya diseñado dicha intervención. La evaluación no toma los objetivos como un dato (como hace la auditoría), sino que también los evalúa. Por ejemplo, puede suceder que el objetivo de un proyecto fuera construir un camino para unir dos ciudades y que ese objetivo se haya alcanzado y de un modo eficiente, pero que dicho camino no fuera prioritario, ya que existía una ruta alternativa, en tanto que la estrategia de desarrollo humano del país daba especial importancia a la mejoría de la salud en zonas rurales a través de la construcción de centros de atención comunitaria. En este caso, la relevancia de la intervención sería en principio muy reducida (nótese que este tipo de análisis puede hacerse aun antes de que se complete la intervención y que no es afectado por problemas de atribución, algo que será considerado más adelante).

La eficacia, o efectividad, es el grado en que se alcanzan los objetivos de una intervención. Como se mencionó anteriormente, si los objetivos contemplan un horizonte temporal que no se limita al corto plazo, el juicio acerca de la eficacia incluye la sostenibilidad de los resultados de la intervención. Frecuentemente se confunde la sostenibilidad de una intervención con la sostenibilidad de sus resultados, y la experiencia y la literatura del desarrollo incluye abundantes referencias a la no sostenibilidad de unidades ejecutoras de proyectos, así como recomendaciones aparentemente lógicas, como no incluir este tipo de unidades en el diseño de intervenciones. En rigor, lo que debería importar es la sostenibilidad de los resultados y, en algunos contextos, como en el caso de las economías que carecen de estructuras sólidas —sea por catástrofes naturales o por encontrarse en una primera fase de transición desde otro tipo de sistema—, dichas unidades ejecutoras pueden jugar el papel del andamio de una construcción. Y así como no tendría sentido que un andamio siguiera montado después de la construcción, lo mismo pasaría con las unidades ejecutoras una vez que se hayan alcanzado los resultados esperados (que pueden incluir el desarrollo de capacidades institucionales). En este marco, la desaparición de las unidades ejecutoras no sería un problema, sino, por el contrario, una

de las condiciones para la eficacia de la intervención. Pero ¿cómo hacer para saber si un determinado objetivo (por ejemplo, el incremento en los ingresos de los campesinos de una región) fue debido a un proyecto de microcrédito o a otros factores? ¿Cómo atribuir los cambios observados en las variables-objetivos a determinadas intervenciones? Este es uno de los problemas más serios que deben enfrentar los evaluadores y volveremos a considerarlo más adelante.

La eficiencia es el tercer criterio evaluativo. Si un proyecto ha permitido alcanzar un resultado relevante pero a un costo muy alto, la evaluación deberá poner en evidencia esta deficiencia. Puede haber otros medios más eficientes para alcanzar los mismos resultados, u otros fines a los cuales pueden, o podrán, asignarse los recursos. Hay diferentes formas en que es posible medir la eficiencia, varias de las cuales son variantes del análisis costo-beneficio. También medidas de costos unitarios pueden ser útiles en algunos contextos. Lo importante es relacionar los logros con los medios empleados para alcanzarlos y elaborar medidas que puedan compararse con las que se verifican en otros usos reales o potenciales del mismo monto y tipo de recursos.

Esquema 1. Relación entre relevancia, eficacia y eficiencia



La relación entre relevancia, eficacia y eficiencia puede mostrarse mediante el diagrama siguiente: la relevancia examina los objetivos a la luz de la estrategia; la eficacia corresponde al grado de logro de los objetivos, y la eficiencia es la relación entre los niveles de logros alcanzados y los recursos utilizados (Esquema 1).

6. El problema de la atribución y algunas formas de encararlo

Antes hemos hecho referencia al problema de la atribución, que vamos a abordar en esta sección. Ante todo, es importante tener en cuenta que hay un conjunto de factores que operan simultánea y sucesivamente en la realidad, en una realidad compleja, y que, por lo tanto, generalmente hay múltiples causas y no causas únicas. En el ejemplo mencionado anteriormente de un proyecto o programa de microcrédito para una población rural (que puede ser un componente clave de una política contra la pobreza), es importante considerar la eventual existencia de otros programas de microcrédito, factores climáticos, políticas que afecten el precio de los insumos y/o productos de los prestatarios y, de esta forma, la rentabilidad de las actividades y, en consecuencia, la capacidad de repago de los microcréditos, etcétera (el etcétera en este caso cubre no solo los factores conocidos, sino también algunos que pueden no haber sido percibidos todavía, pero que en el futuro pueden revelarse como importantes; por ejemplo, durante bastante tiempo no se prestó suficiente importancia a los costos de transacciones y a los factores que los afectan, y hoy hay consenso en que dichos costos son un factor importante en el resultado de los programas de microcrédito).

Esquema 2. Múltiples factores causales



El Esquema 2 muestra el efecto de varios factores, incluida una intervención específica que puede ser una política pública (o programa o proyecto), sobre un conjunto de variables clave (aquellas que interesa influenciar, a veces llamadas variables de resultado).

Es útil plantear cuatro preguntas relacionadas:

- a) ¿Cómo identificar el efecto de la intervención X en las variables clave?
- b) ¿Cómo aislar o separar dicho efecto del correspondiente a otros factores?

- c) ¿Qué hubiera sucedido en ausencia de la intervención X?
- d) ¿Qué se suponía que hubiera sucedido si no se realizaba la intervención X?⁶.

Las preguntas precedentes son útiles para abordar el tema de la atribución. La pregunta fundamental es la tercera, ¿qué hubiera sucedido en ausencia de la intervención X? Frecuentemente, de modo implícito o explícito, se hacen supuestos sobre lo que sucedería si no se realiza dicha intervención (lo cual precisamente se usa como justificación para la intervención y constituye el contrafactual o contrafáctico, implícito o explícito)⁷.

Las primeras dos preguntas requieren un trabajo analítico, de separación de efectos e identificación de causas, que a veces (especialmente en el campo de la evaluación de la cooperación internacional) se tiende a considerar imposible o de interés solo académico. Se sostiene que, en lugar de ocuparse y preocuparse en la atribución de resultados a determinadas intervenciones, lo importante es establecer la contribución de determinada intervención a los resultados. En rigor, esto no es una solución, sino un desplazamiento del problema, ya que implica que hay que identificar la contribución, distinguiéndola de la contribución de otros efectos⁸, con lo cual el problema continúa. ¿Cómo abordarlo?

Una forma de hacerlo es a través de la utilización de grupos de control, de un diseño de la evaluación que permita establecer comparaciones entre dos grupos, el de aquellos que han sido objeto de la intervención y otros cuyos integrantes difieren del primero solo en que no han sido afectados por la intervención⁹. De poder establecerse un grupo de control, algo que discutiremos más adelante, podrán realizarse comparaciones sobre las diferencias antes y después de la intervención y entre quienes fueron y los que no fueron objeto de esta. Se tienen así cuatro casos que corresponden a la situación antes y después de la intervención, y con/sin intervención (Esquema 3).

Esquema 3. Grupos con y sin intervención, antes y después de la intervención

| | Con intervención | Sin intervención |
|---------------|------------------|------------------|
| Antes | CA | SA |
| Después | CD | SD |

La comparación antes y después limitada a la población que fue objeto de la intervención (CA)/(CD) no permite tomar en cuenta el efecto de otros factores. En cambio, si se analizan las diferencias (CA)/(CD) y (SA)/(SD), se podrían comparar los cambios antes y después de la intervención para el grupo afectado por la intervención, con los cambios antes y después experimentados por el grupo no afectado por la intervención. A modo de ejemplo, tomemos un caso donde la variable medida sea el ingreso medio del grupo antes de un proyecto de microcrédito (Esquema 4).

Esquema 4. Ejemplo de una intervención que ha permitido mantener el ingreso*

| | Con intervención | Sin intervención |
|---------------|------------------|------------------|
| Antes | 100 | 100 |
| Después | 100 | 80 |

Nota: *Unidad monetaria \$.

En este ejemplo, si se comparara la situación antes y después de quienes fueron objeto de la intervención, se observaría que no hubo cambio en el ingreso y esto podría ser interpretado como un fracaso del proyecto. Sin embargo, la comparación con el grupo de control muestra que, de no haber sido por el proyecto, los ingresos hubieran caído en un 20 por 100, pasando de 100 a 80 unidades monetarias. Una situación simétrica se presentaría si los ingresos del grupo bajo el proyecto aumentan en un 20 por 100, en tanto que los ingresos de la población fuera del proyecto tienen un incremento de ingresos del 40 por 100, en cuyo caso lo que parece un éxito sería en realidad un fracaso. Aunque no siempre, a veces las apariencias engañan. Y los grupos de control evitan caer en el engaño.

La pregunta que inmediatamente surge es: ¿resulta prácticamente posible realizar evaluaciones con grupos de control? ¿O se trata de una posibilidad meramente teórica? La respuesta breve, que se desarrollará a

continuación, es que, en algunos casos y para cierto tipo de intervenciones, es posible y conveniente realizar una evaluación con grupos de control. Pero que hay numerosas e importantes situaciones en que no es posible aplicar este enfoque.

Los tipos de intervenciones en que los grupos de control son factibles y recomendables consisten en intervenciones muy específicas (estandarizadas), que no varían durante el periodo de implementación, y que se aplican a un subconjunto de la población. Hay casos en los cuales la implementación se realiza usando procedimientos aleatorios (loterías) para racionar recursos limitados, lo cual es frecuentemente bien recibido por las poblaciones como un mecanismo transparente de asignación de recursos. Y, cuando esto sucede, las condiciones son particularmente propicias para realizar evaluaciones con grupos de control¹⁰. Además, se están haciendo cada vez más esfuerzos para incorporar, durante la fase de diseño de los proyectos y programas, sistemas de seguimiento y evaluación dotados de fondos suficientes y desarrollados por gente con la pericia necesaria para asegurar que dichos sistemas procedan a la recopilación de datos durante el periodo de implementación que haga posible una evaluación con grupos de control. Esto se ha debido en parte a la presión por generar evidencias rigurosas de resultados alcanzados por la cooperación al desarrollo (y por programas y proyectos financiados con recursos presupuestarios propios), así como al desarrollo de capacidades para realizar este tipo de evaluaciones.

Pero este tipo de evaluación no es factible en el caso de evaluaciones de políticas globales, como una política monetaria, o de proyectos o programas en los cuales las intervenciones se van modificando durante la implementación, o en el caso de programas nacionales en los cuales toda la población sea objeto del tratamiento y la estrategia de implementación no pueda incorporar la aleatorización como procedimiento para asignar los recursos. Sin embargo, existen algunos métodos que también se han aplicado para la evaluación rigurosa de estos últimos tipos de programa¹¹.

Es importante que los evaluadores tengan presente que puede haber hipótesis rivales plausibles que expliquen los resultados alcanzados (o no alcanzados), como el efecto de precios no influenciados por la intervención, o migraciones, u otros factores externos, incluyendo otras intervenciones.

Uno de los desafíos principales que enfrentarán los evaluadores es identificar dichas hipótesis rivales plausibles y considerar hasta qué punto es posible descartarlas. Para ello es útil proceder a un análisis comparativo, efectuando comparaciones con situaciones donde esos mismos factores externos hubieran estado operando, aun si no es posible contar con grupos de control. Vale la pena mencionar que un aspecto importante de la experiencia de desarrollo de China es la experimentación de distintos sistemas de producción y distribución, incluyendo tecnologías de diferentes características y un proceso de evaluación no formal que identifica aquellos experimentos que dieron mejor resultado y que son reproducidos a escala más amplia en el resto del país, difundiendo de esta forma los resultados del aprendizaje¹².

Finalmente, cabe destacar que, a pesar de una tendencia reciente a privilegiar resultados y desvalorizar la importancia de los procesos, las dificultades que presenta la atribución de resultados lleva a que sea conveniente tomar en cuenta adecuadamente los procesos asociados a una intervención, ya que estos pueden aportar una clarificación de la atribución¹³. Consiguientemente, las evaluaciones de procesos pueden complementar las evaluaciones de resultados.

7. Enfoques para la evaluación de políticas públicas

En la práctica de la evaluación, hay una distinción central en la forma en que las políticas públicas son concebidas: por un lado, como tratamientos, cuya unidad de referencia son programas o proyectos; por otro lado, las políticas son conceptualizadas como intervenciones que plantean un marco normativo, que no siempre puede reducirse a programas o proyectos. Los métodos que se emplean o pueden emplearse para la evaluación en ambos casos no son necesariamente los mismos, tal como se expone en el resto de este apartado.

Políticas como «tratamientos»

El enfoque de las políticas como tratamientos, que aborda el problema de la evaluación de políticas del mismo modo en que se encara la evaluación de fármacos, fue considerado en textos clásicos de evaluación, como

en Campbell y Stanley (1963), y ha resurgido entre los economistas, incluyendo un premio Nobel, James Heckman, cuyo discurso de aceptación del premio fue precisamente sobre la evaluación de políticas públicas¹⁴. La literatura sobre los efectos de tratamientos investiga una clase de intervenciones con cobertura parcial (que no incluyen a toda la población, o sea, que no son universales), de modo tal que hay un grupo de tratamiento y otro de control, y se reconoce que no es útil para evaluar intervenciones aplicadas universalmente en una economía (o sociedad); supone que los efectos en los miembros del grupo de control son los que resultarían en ausencia de una intervención.

Por ejemplo, la evaluación de políticas activas en el mercado de trabajo utilizando este enfoque ha mostrado que a lo largo del tiempo, y en varios países, este tipo de políticas han sido ineficaces en promover el crecimiento de largo plazo del empleo y los salarios. Se ha encontrado que las comparaciones de los resultados postprogramas de personas que participan en programas de formación con las medidas preprogramas para dichas personas (procedimiento utilizado frecuentemente) sobreestiman sustancialmente las ganancias derivadas de la participación en la formación ya que buena parte de la mejora postprograma hubiera ocurrido en ausencia de los programas de capacitación¹⁵.

Si bien este enfoque ha producido resultados importantes, su campo de aplicación es limitado a intervenciones que se apliquen de modo estandarizado, sin variaciones durante el periodo de implementación¹⁶. Además, en el marco de las reformas del sector público, los hacedores de política (los *policy-makers*) hacen preguntas sobre políticas y estrategias más que sobre programas o proyectos. Como señala Elliot Stern, muchos programas buenos no son equivalentes a una buena política¹⁷. Por eso es necesario considerar un enfoque alternativo, que es el tema de los próximos párrafos.

Políticas públicas como intervenciones no reducibles a programas o proyectos

Si bien las políticas públicas frecuentemente se traducen en programas y proyectos, es importante observar que la evaluación de un conjunto de

programas y proyectos vinculados a una política pública no equivale a una evaluación de dicha política.

En efecto, sucede a veces que los programas y proyectos alcanzan sus objetivos de modo eficiente, y dichos objetivos son relevantes, pero algunos de los aspectos del marco normativo que constituye la política en cuestión pueden no haberse concretado en absoluto. Por ejemplo, en el caso de una política educativa que plantee el acceso universal a la educación primaria y que cuente con un conjunto de programas y proyectos innovativos que prueban nuevas técnicas pedagógicas de bajo costo y alta calidad utilizando nuevas tecnologías de comunicación, aun si dichos programas alcanzan plenamente su objetivo de mostrar que dichas técnicas funcionan conforme a las expectativas, la política educativa puede sin embargo fracasar si no hay una adecuada replicación (o masificación) de esas técnicas. O sea, que la política podría fracasar aun si los programas y proyectos con ella relacionados hubieran sido exitosos.

Para evaluar las políticas públicas, es importante primero aclarar su marco orientativo o normativo¹⁸, tratando de identificar los objetivos planteados explícitamente. Cuando no se encuentren dichos objetivos se deberá realizar una reconstrucción (o interpretación) de estos ya que, de lo contrario, no sería posible evaluar la política. Una vez que se clarifiquen los objetivos de la política pública, pueden aplicarse los criterios de evaluación tal como fueron formulados más arriba, eventualmente expandiéndolos para incorporar algún(os) criterio(s) que resulte relevante (como el ya mencionado de equidad).

Además de aplicar los criterios de evaluación típicos (relevancia, eficacia y eficiencia), también es útil, cuando se evalúan políticas públicas, considerar su efecto sobre los incentivos que generan (y hasta qué punto estos son necesarios o suficientes para que se alcancen los objetivos), las capacidades necesarias para implementar las políticas así como el efecto de dichas políticas sobre las capacidades, la adopción de nuevos comportamientos o prácticas o tecnologías promovidas por la política pública, los riesgos asociados a la implementación de la política (particularmente la posibilidad de que se generen efectos no buscados) y, finalmente, las oportunidades creadas por la política pública (la equidad puede en este contexto ser un

criterio importante ya que llevaría a considerar el grado en que la política pública ha promovido o no la igualdad de oportunidades)¹⁹.

La discusión precedente puede resumirse en el Esquema 5, que incorpora además un par de cuestiones adicionales:

- La distinción entre la evaluación independiente, realizada por quienes no han participado en el diseño o la implementación de una determinada política, y la autoevaluación, llevada a cabo por quienes han participado en su diseño o implementación. Las mejores prácticas evaluativas combinan ambos tipos de evaluaciones, de modo tal que la evaluación independiente revisa la autoevaluación y la toma como un insumo que permite aumentar su cobertura.
- La necesidad de considerar los factores que influyen sobre el uso de las evaluaciones, destacándose tres de ellos: la oportunidad (la evaluación debe estar disponible cuando se tomen decisiones), la credibilidad (que no se ponga en duda la objetividad y solidez de la evaluación) y su diseminación o difusión (que las evaluaciones sean conocidas, y no archivadas como un documento histórico). Una precondition para este último factor es la transparencia del proceso evaluativo (la posibilidad de colocar las evaluaciones en Internet es una forma de facilitar la transparencia).

Esquema 5. Resumen de la evaluación de políticas públicas

| | | |
|---------------------|---------------------------------|---|
| 1. Fines | ¿Por qué evaluar? | Aprendizaje Rendición de cuentas |
| 2. Agentes | ¿Quién evalúa? | Autoevaluación Evaluación independiente |
| 3. Objeto | ¿Qué evaluar? | Proyectos, programas, políticas públicas |
| 4. Momento | ¿Cuándo evaluar? | Ex ante, tiempo real, ex post |
| 5. Cobertura | ¿Cuánto evaluar? | Costos y beneficios Nivel óptimo |
| 6. Métodos | ¿Cómo evaluar? | «Tratamientos» Criterios: relevancia, eficacia, eficiencia |
| 7. Uso | Factores que influyen en el uso | Oportunidad, credibilidad, diseminación |

8. Institucionalización de la evaluación

La evaluación como disciplina se ha ido consolidando en Europa, a partir de la década de los noventa, con la creación en 1995 de la Sociedad Europea de Evaluación (EES), y con el surgimiento, catalizado por la EES, de asociaciones nacionales de evaluación en varios países europeos, incluyendo España, donde en el año 2001 se creó la Sociedad Española de Evaluación de Políticas Públicas (SEE)²⁰.

Cabe también destacar las experiencias de institucionalización de la evaluación en países de América Latina. La más destacada y evaluada es la experiencia de Chile, que ha constituido un sistema de evaluación con una orientación de resultados²¹.

9. Dificultades y soluciones

En esta sección se abordarán algunas de las dificultades más importantes para que la evaluación se constituya en un instrumento de aprendizaje, y para cada una de dichas dificultades se sugerirán soluciones. Se destaca la importancia de emplear una metodología transparente, clarificando y difundiendo los criterios que se utilizarán en las evaluaciones. También se enfatiza la necesidad de llevar a cabo acciones que hagan más probable que las evaluaciones sean utilizadas. Además, se señala la conveniencia de fomentar la realización de autoevaluaciones y su revisión por entes independientes, configurando una arquitectura de la evaluación, a varios niveles, que corresponda a una cultura de la evaluación.

Por otra parte, se plantea la importancia de desarrollar capacidades de evaluación, tanto a través de mecanismos formales (facilitando el acceso a estos), como mediante la creación de oportunidades para «aprender haciendo», involucrando a universidades y centros de investigación que tienen una capacidad real o potencial para realizar evaluaciones. En este sentido, se discute el papel que puede tener un fondo para la financiación de evaluaciones. Finalmente, se sugiere la conveniencia de introducir incentivos positivos para que se realicen evaluaciones de calidad, tales como premios anuales a las mejores evaluaciones, lo cual también puede contribuir al desarrollo de una cultura de la evaluación:

1. *Falta de fondos*: frecuentemente, este suele ser mencionado como el principal problema o dificultad para la realización de evaluaciones y para poder disponer de un acervo o *stock* que constituya un repositorio de conocimiento útil como fuente de aprendizaje. Si bien esto es —en general— cierto, también se ha observado en varias ocasiones que fondos asignados para evaluaciones (lo cual constituye una forma de hacer frente a esta dificultad, estableciendo por ejemplo que un porcentaje del 2 por 100 de los fondos de la intervención se dediquen a la evaluación) no se utilizan o se reasignan para otras finalidades, por las razones que se mencionan a continuación.
2. *Capacidades limitadas*: generalmente, suele haber una brecha significativa entre la demanda de instituciones y expertos en evaluación y la oferta existente, particularmente dada la fuerte expansión de la demanda motivada por la presión de mostrar resultados, así como por las restricciones de fondos y la búsqueda de un mejor uso de estos. Resulta útil distinguir entre la oferta actual o real de evaluadores, incluyendo instituciones, y la oferta potencial, incluyendo en esta a universidades y centros de investigación y a profesionales sin experiencia en evaluación, pero que podrían ser motivados para que desarrollen su capacidad en este campo a través de la oportunidad de «aprender haciendo», lo que complementaría la formación que puede realizarse mediante cursos y seminarios. Un programa de becas para tesis y tesinas sobre evaluación financiado, por ejemplo, por fundaciones vinculadas a bancos u otras organizaciones, públicas y/o privadas, podría favorecer el desarrollo de capacidades de evaluación.
3. *Cuestiones metodológicas*: prácticamente, todos los métodos tienen limitaciones y es importante desarrollar un enfoque pragmático, empleando diferentes métodos, triangulando tanto las fuentes de información como los métodos que se utilizan. Además, una dificultad frecuente es que las evaluaciones son mucho menos utilizadas de lo que se espera, y esto se debe en gran medida a que no se presta atención suficiente al uso de las evaluaciones. Es útil recurrir, en la medida de lo posible, a expertos en comunicación que puedan ayudar a presentar las evaluaciones de modo que resulten más «digeribles» (o «amigables»), aplicando un principio

de división del trabajo, pero también es importante que los evaluadores sean responsables de participar en el proceso de difusión de las evaluaciones y no solamente en su producción. Los esfuerzos que se hagan en la fase de postproducción de las evaluaciones y la calidad de dichos esfuerzos son fundamentales para que las evaluaciones sean utilizadas como instrumento de aprendizaje²². Cabe también destacar que, en la medida en que las evaluaciones cuestionen los supuestos básicos de las intervenciones, podrán también facilitar un proceso de *double-loop learning*²³, o macroaprendizaje, que permita extraer lecciones que vayan más allá de la intervención concreta que se evalúe.

4. *Falta de incentivos*: la evaluación suele ser percibida como una amenaza. Esto ha llevado a que, en algunos casos, se reemplace el término evaluación por el de revisión, lo que no resuelve el problema; la dificultad tiene que ver, en parte, con el uso anticipado de la evaluación para la rendición de cuentas, que suele enfatizarse a expensas del uso de la evaluación como instrumento de aprendizaje. A menudo se plantea que hay una opción, un *trade-off*, entre estas dos funciones de la evaluación: o se privilegia la rendición de cuentas o el aprendizaje. Una perspectiva alternativa que tiene más sentido es que la función de rendición de cuentas complementa el papel de la evaluación como instrumento de aprendizaje, ya que la rendición de cuentas puede constituirse en un incentivo para aprender puesto que se trata también de rendir cuentas del aprendizaje realizado. Lo que es importante es desarrollar una cultura de la evaluación, en la cual los errores sean percibidos como oportunidades para el aprendizaje, y el aprendizaje como una fuente importante del crecimiento y desarrollo. La instauración de premios para las mejores evaluaciones pueden ser incentivos para mejorar su calidad. Pero, en última instancia, los incentivos principales para que se realicen las evaluaciones y para que estas sean de calidad dependen de la demanda de evaluaciones que hagan los políticos y la población²⁴. La producción y adecuada difusión de evaluaciones de calidad puede así iniciar un círculo virtuoso, fortaleciendo la demanda por evaluaciones.

10. Síntesis y conclusiones

La evaluación pragmática de políticas públicas no se limita a preconizar la utilización de un solo método o enfoque; rechaza el fundamentalismo metodológico buscando, en cambio, el método adecuado a las circunstancias, alineándose con el principio del rigor apropiado. También es pragmática porque enfatiza el uso de la evaluación, sin darlo por descontado y adoptando un papel activo para promover la difusión y aplicación de las evaluaciones de políticas públicas. La discusión precedente puede sintetizarse en el siguiente conjunto de diez conclusiones:

- La evaluación bien realizada y adecuadamente utilizada permite aprender de la experiencia, extrayendo lecciones sobre las políticas que han funcionado bien y las que han fallado, y las razones de éxitos y fracasos, propios y ajenos.
- Los criterios de evaluación (relevancia, efectividad y eficiencia, y algunos otros como equidad, que pueden incluirse en función de las demandas de los políticos) permiten considerar sistemáticamente (y extraer lecciones de) diversos tipos de políticas públicas.
- Dado que las políticas públicas que se evalúan se implementan en un contexto complejo en el que operan múltiples factores, es importante evitar un enfoque ingenuo de la atribución que no reconoce la existencia de otros factores que no sea la propia política pública que se evalúa.
- Para políticas públicas en las que los grupos de control o de comparación sean factibles, particularmente aquellas que consisten o se componen de intervenciones muy específicas (programas o proyectos) que no varían durante el periodo de implementación y que se aplican a un subconjunto de la población, es recomendable realizar la evaluación utilizando dichos grupos de control.
- Cuando no sea factible utilizar grupos de control, es importante considerar hipótesis rivales plausibles, y utilizar evidencias para determinar cuál es la hipótesis más sólida. Además, sobre todo en estos casos en que la atribución es más conjetural, es importante

complementar las evaluaciones de resultados con evaluaciones de procesos.

- Con el propósito de hacer frente a la falta de fondos para las evaluaciones, es útil considerar la creación de un fondo para evaluaciones, que puede constituirse con un pequeño porcentaje del total de fondos dedicados a políticas y programas públicos (por ejemplo, aproximadamente un 2 por 100). Parte de estos fondos podrían utilizarse para autoevaluaciones y parte para evaluaciones independientes que pueden centralizarse en una unidad o agencia de evaluación, la cual podría reducir los costos de las evaluaciones (por economías que resultan de realizar tareas similares en diferentes contextos, economías of scope, y por economías de escala) y facilitaría el diseño de evaluaciones que permitan análisis comparativos y metaanálisis así como metaevaluaciones.
- La formación en evaluación puede complementarse mediante la creación de oportunidades para «aprender haciendo», viabilizadas por un programa de becas para tesis y tesinas sobre evaluación, financiado, por ejemplo, por fundaciones vinculadas a bancos u otras organizaciones privadas o estatales.
- Es importante desarrollar un enfoque metodológico pragmático, utilizando diferentes métodos para realizar las evaluaciones en función de los tipos de políticas públicas y de la información disponible.
- Los evaluadores y los gestores de evaluaciones deben ocuparse no solamente de la producción, sino también de la postproducción de evaluaciones, asumiendo una responsabilidad por facilitar su utilización. La división de trabajo que involucre a expertos en comunicación puede ser útil para que las evaluaciones no se limiten a ser solamente informes, sino que se conviertan en instrumentos valiosos para el aprendizaje.
- La instauración de premios para las mejores evaluaciones o para las evaluaciones más influyentes pueden ser incentivos para mejorar su calidad. Pero los incentivos principales para que las evaluaciones sirvan como instrumento de aprendizaje están dados por la demanda de evaluación que hagan los políticos y la población.

Notas

* Título original “La evaluación pragmática de políticas públicas”, publicado en Información Comercial Española, Revista ICE, Madrid, Mayo-Junio 2007, N.º 836, pp. 9-31. Reproducción autorizada.

** Consultor Independiente, ex Gerente y Asesor del Departamento de Evaluación del Banco Mundial. Versión de noviembre de 2006.

1. A veces (particularmente en el caso de proyectos o programas), se diferencia entre la evaluación terminal, que se realiza dentro de los seis meses de concluida una intervención, y la evaluación *ex post*, que se lleva a cabo después de algunos años (por ejemplo, 3) después de finalizada la intervención. Ambas son evaluaciones a *posteriori*.
2. Sobre la relación entre políticas públicas y política, véase Franco y Lanzaro (2006), especialmente el capítulo 1 sobre determinación y autonomía de las políticas públicas.
3. Los dos primeros criterios adicionales mencionados en este párrafo son utilizados por el departamento de evaluación del Banco Mundial y la mayor parte de los bancos multilaterales de desarrollo, en tanto que el tercero es de uso frecuente en Europa.
4. En algunos contextos puede ser importante enfatizar la importancia de la sostenibilidad y, para ello, puede resultar conveniente considerarla como un criterio adicional.
5. Por ejemplo, la consideración de la equidad como un criterio evaluativo permite abordar de modo más sistemático cuestiones de género y exclusión.
6. Notar que la pregunta c) se refiere a lo que hubiera sucedido, a juicio del evaluador, si la intervención X no se hubiera realizado, en tanto que la d) corresponde a lo que se suponía en el diseño de la intervención que hubiera sucedido en ausencia de X. El evaluador puede no estar de acuerdo con el supuesto contrafactual del diseño de la intervención y, por eso, es conveniente efectuar la distinción. Suele suceder que el diseño de la intervención no explicita el supuesto contrafactual, pero generalmente hay por lo menos un supuesto implícito.
7. Para una discusión sobre los contrafácticos, veáse por ejemplo Elster (2006), cap. 6.
8. Véase Feinstein (2006).
9. Ver, por ejemplo, Weiss, C. (1982).
10. Como ejemplos véanse los trabajos citados en Pitman, G. K. *et al.* (2005) y Weiss, C. (1982).
11. Sobre la limitación de la aleatorización, y sobre algunos métodos alternativos, véase Pitman, G. K. *et al.* (2005).
12. Este proceso de evaluación, difusión y aprendizaje no parece estar todavía documentado adecuadamente.

- ¹³. Es interesante establecer una analogía con lo que ha sucedido en economía, disciplina en la cual a comienzos de los años cincuenta se colocó el énfasis en las predicciones con cierto menosprecio por los supuestos, hasta que comenzó a percibirse que desde un punto de vista práctico era conveniente volver a dar importancia a los supuestos, y no solo a las predicciones, cuando se juzgan las teorías económicas
- ¹⁴. Ver Heckman (2001); este artículo de 75 páginas tiene una bibliografía de 11 páginas de trabajos microeconómicos de evaluación con el enfoque del «tratamiento». Una breve discusión de Heckman sobre cuestiones de evaluación puede encontrarse en Feinstein y Picciotto (2001), páginas 83-87.
- ¹⁵. Heckman (2001), páginas 713, 730 y 732.
- ¹⁶. Ver Feinstein (2006) y Pitman, Feinstein e Ingram (2005).
- ¹⁷. Stern (2006), página 299 *many good programmes do not make a good policy*.
- ¹⁸. Por marco normativo en este contexto se refiere a normas que no necesariamente requieren un marco legal, sino que constituyen pautas para orientar las intervenciones
- ¹⁹. Estas cinco dimensiones pueden recordarse fácilmente utilizando un recurso mnemotécnico: las iniciales de las palabras subrayadas componen el nombre de ÍCARO, el personaje de la mitología griega que trató de volar hacia el sol.
- ²⁰. Para un panorama de la evaluación en Europa y en el mundo, véase el Atlas Internacional de la Evaluación, Furubo, J. E., Rist, R. C. y Sandahl, R. (2002), que incluye un capítulo sobre España.
- ²¹. Véase Rojas, F. *et al.* (2006).
- ²². Los factores que afectan el uso de las evaluaciones son analizados en Feinstein (2002).
- ²³. Para una discusión del *double-loop learning* en el marco del aprendizaje organizacional, véase Leeuw *et al.* (1993).
- ²⁴. Ver, por ejemplo, Ramiro Sánchez (2002).

Referencias

- Bandura, A. (1982): *Teoría del Aprendizaje Social*, Espasa Calpe, Madrid.
- Elster, J. (2006): *Lógica y Sociedad*, Gedisa, Barcelona.
- Feinstein, O. (2002): «Use of Evaluations and the Evaluation of their Use», *Evaluation*, 8, 4, 433-439.
- Feinstein, O. (2006): «Evaluation of Development Interventions and Humanitarian Action», con Tony Beck, en Shaw, I. *et al.* (eds.): *Handbook of Evaluation*, Sage, Londres.

Feinstein, O. y Picciotto, R. (eds.) (2001): *Evaluation and Poverty Reduction*, Transaction, New Brunswick.

Franco, R. y lanzaro, J. (2006): *Política y Políticas públicas en los procesos de reforma en América Latina*, FLACSO/Miño y Dávila, Buenos Aires.

Heckman, J. (2001): «Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture», *Journal of Political Economy*, volumen 100, número 4, 2001, páginas 673-748.

Heckman, J. (2001) en Feinstein, O. y Picciotto, R. (eds.) (2001): *Evaluation and Poverty Reduction*, Transaction, New Jersey, páginas 83-87.

Leeuw, F.; Rist, R. y Sonnichsen, R. C. (eds.) (1993): *Can Governments Learn?*, Transaction, New Brunswick.

Pitman, G. K.; Feinstein, O. E Ingram, G. K. (eds.) (2005): *Evaluating Development Effectiveness: Challenges and the Way Forward*, Transaction Publishers, New Brunswick.

Ramiro Sánchez, A. (2002): *Demandas de calidad de la Administración Pública: Un derecho de la ciudadanía*, Dykinson, Madrid.

Rojas, F.; Mackay, K.; Matsuda, Y.; Shepherd, G.; Del Villar, A.; Zaltsman, A. y Krause, P. (2006): *Chile: Análisis del Programa de Evaluación del Gasto Público*, Banco Mundial, Washington D.C.

Stern, E. (2006): *Contextual Challenges for Evaluation Practice*, en Shaw, I. et al. (2006).

Weiss, C. (1982): *La investigación evaluativa*, Trillas, México.

Dos métodos de evaluación: criterios y teoría del programa*

Juan Andrés Ligeró Lasa

1. Introducción

No hay una única forma de evaluar. En un primer estadio de la evaluación que se puede fijar en torno a los años 50 del siglo xx, quizá sí existía un amplio acuerdo sobre el método a utilizar, el experimental: los diferentes actores esperaban validar los programas en función del cumplimiento de los objetivos declarados. La evaluación aportaba confianza y optimismo en sus posibilidades. Pero los resultados de las evaluaciones no llegaban a las expectativas depositadas, había dificultades prácticas, críticas metodológicas, problemas técnicos y resultados no concluyentes; algo que alejaba el ideal de convertir la evaluación en el apoyo científico de la decisión política.

Después de cinco décadas de realismo en cuanto al uso y las posibilidades de la evaluación, han emergido diferentes métodos, reflexiones y aproximaciones sobre cómo hacer evaluación. Por tanto, existen diversos abordajes, con diversos autores y escuelas que piensan y trabajan de forma distinta. Stufflebeam (2001) los ha clasificado en al menos 22 grandes corrientes. Esto quiere decir que hacer evaluación también implica decidir sobre el procedimiento a utilizar entre una amplia gama de opciones.

En este contexto teórico, llama la atención que en el sector de la cooperación para el desarrollo se dé la presencia hegemónica de un determinado método de evaluación, como es el *modelo preordenado, por criterios y estándares*

propuesto por el Comité de Ayuda al Desarrollo (CAD) de la Organización para la Cooperación y el Desarrollo Económico (OCDE).

El presente texto pretende revisar el modelo en sus virtudes y defectos, y proponer al menos otra forma alternativa de proceder, la *evaluación de la teoría del programa*. No deja de ser una limitación el proponer solamente una única alternativa sabiendo que existen otras aproximaciones lúcidas y útiles. Pero el objetivo del texto es mostrar que hay otros métodos posibles y no aspirar a convertirse en una recopilación de teorías y modelos, algo que por otro lado ya está escrito.

2. Evaluación orientada por criterios

Un criterio es una definición sobre la situación deseable u óptima (Bustelo, 2010) de un proyecto en una determinada parcela de la realidad. De forma resumida se puede decir que la evaluación por criterios funciona sometiendo a examen al programa y valorándolo con relación a ciertos criterios establecidos. Para poder emitir un juicio sobre el programa, se requiere establecer unos puntos críticos en cada uno de los criterios, a los que se los denomina «estándares». Si se alcanza el estándar, se considera logrado el criterio. Por ejemplo, en el criterio eficiencia se marca como estándar que los beneficios superen los recursos invertidos. Si la evaluación concluye que esto es así, el programa se considera eficiente.

Para diseñar una evaluación por criterios, se proponen las siguientes fases¹:

Fases de la evaluación orientada por criterios:

- 2.1. Elección entre modelos estandarizados o modelos *ad hoc*.
- 2.2. Identificación de fuentes y actores clave.
- 2.3. Extracción de las necesidades informativas
 - a) Fase divergente
 - b) Fase convergente
- 2.4. Operacionalización
 - a) Trabajo vertical. Definición de los criterios y preguntas.
 - b) Trabajo horizontal. Definición de indicadores y fuentes.

2.1. Elección entre modelos preordenados o modelos *ad hoc* sensibles a las necesidades de los actores implicados

La evaluación preordenada (también llamada *estandarizada*) es aquella donde las cuestiones a evaluar están previamente establecidas. Por el contrario, la evaluación *ad hoc* o sensible a las necesidades de los diferentes actores se construye identificando los valores, ya sean explícitos o subyacentes, de los diferentes agentes implicados. Es en el propio proceso de evaluación donde emergen los criterios o las preguntas que se consideran relevantes evaluar.

El éxito de los modelos preordenados probablemente radique en utilizar un mismo formato para evaluar las intervenciones, sea cual sea la evaluadora² y esté donde esté la acción. Unifican el lenguaje y disipan incertidumbres sobre cómo acometer el trabajo, sobre todo en las primeras fases, ya que solo hay que disponer los criterios preestablecidos y los mecanismos que proponga el método. Tal es el caso del modelo propuesto por el CAD de la OCDE (1995). El CAD plantea cinco criterios para evaluar las acciones de cooperación para el desarrollo: impacto, eficacia, eficiencia, pertinencia y sostenibilidad. Esta propuesta ha tenido una gran aceptación y ha sido asumida por los principales actores de la cooperación internacional. En España, la Secretaría de Estado para Cooperación Internacional del Ministerio de Asuntos Exteriores y de Cooperación la desarrolló y difundió a través de su manual *Metodología de Evaluación de la Cooperación Española* (1998).

La idea de una estructura única de evaluación ha generado críticas. Lo considerado bueno u óptimo, ya sea en cooperación o en cualquier otro ámbito, no es universal, depende del contexto y de los actores implicados en esa valoración.

En el campo de la acción pública, existen diversos actores con condiciones y posiciones heterogéneas, lo que provoca diferentes perspectivas sobre el programa o sobre lo que debe ser evaluado. Lo que valora como positivo un político sobre un sistema de salud puede que no sea relevante para el beneficiario y puede también ser diferente a lo que esperan los trabajadores. Por tanto, los actores implicados en un programa pueden

tener diferentes expectativas y deseabilidad hacia lo que ofrece una intervención. No se puede presuponer una igualdad de valores entre los diversos actores, ya que ni los contextos sociales ni las condiciones que soportan son homogéneos.

A esta misma conclusión han llegado las metodologías que se conocen bajo el nombre de *calidad*. En un primer momento, a mediados del siglo xx, se entendía la calidad como la tenencia de un conjunto de características de un bien o servicio que se ajustaban a una descripción predeterminada. Es lo que Bouckaert ha denominado la *perspectiva orientada por el productor* (Löffler, 1996: 23). En los años 50 y 60, el concepto de calidad cambia. Se reconoce la existencia de un mercado fragmentado con diversos clientes con sus respectivas necesidades y que además no tienen por qué coincidir. La calidad pasa a ser una función de las expectativas y percepciones individuales (pág. 24). Por esta razón se ha denominado *calidad subjetiva*.

El concepto de calidad ha transitado desde una visión única, la del productor del servicio, hacia las percepciones que de ella tienen los diferentes clientes. Este cambio de enfoque no ha supuesto ninguna pérdida de operatividad o bloqueo metodológico, sino más bien una ganancia al producirse un análisis más realista del sistema y de los actores implicados.

En evaluación ha ocurrido lo mismo, se ha ido introduciendo una perspectiva que también podría llamarse subjetivista en la medida que atiende a sus diversos «clientes». Cronbach argumenta que la evaluación que permite a un solo grupo definir las cuestiones a indagar tiende a no ser útil, ya que la toma de decisiones en sociedades democráticas está caracterizada por el pluralismo y la evaluación debe tener en cuenta este aspecto (Chen, 1990: 61). Aceptar estas premisas supone preguntarse qué actores están implicados en los programas, quiénes tienen legitimidad para participar en un proceso de evaluación y cómo hacer que las evaluaciones respondan a las diferentes necesidades para que sean útiles.

La orientación hacia los actores se la ha llamado *pluralismo* o *responsiveness*³ (Stake, 2006), que podría ser traducido como evaluación sensible a las necesidades de los diversos actores implicados con el programa. Este enfoque recoge una corriente de pensamiento con diversos y significativos

autores como Cronbach, Patton, Guba, Lincoln, Weiss, House y Monnier, entre otros. Pero también ha calado en otras formas de evaluación más «objetivistas» que empiezan a tener cierta empatía hacia los requerimientos de los *stakeholders*⁴. Rossi y Chen defienden un enfoque que integre ambos métodos (Chen, 1990). El mismo CAD describe que se pueden descartar o añadir nuevos criterios adicionales dependiendo de las preguntas evaluativas y de los objetivos de la evaluación (CAD, 2010: 9). En el ámbito nacional, la DGPOLDE⁵ también reconoce que puede haber situaciones donde «se necesite recurrir a criterios complementarios», incluso propone un listado de ellos (MAE, 2007: 57).

Aunque se opte por modelos preordenados de criterios, actualmente hay cierta flexibilidad para adaptarse a los diversos contextos y esquemas valorativos de los diversos actores.

2.2. Identificación de fuentes y actores clave

Como se ha visto en el apartado anterior, en un sistema de acción, como es un programa, existen diferentes *stakeholders* que podrán tener diferentes esquemas y valores (Stake, 2006). La dificultad a la hora de diseñar una evaluación estriba en identificar las fuentes a las que se les otorga legitimidad para participar en el proceso, detectar sus sistemas de valores y necesidades informativas. Eric Monnier (1992) propone tener en cuenta tres subsistemas de actores:

- *Subsistema de la legitimación*: se compone de actores clave a los que denomina responsables legítimos de la decisión. Esta legitimidad puede ser otorgada de forma jurídica, política o científico-técnica.
- *Subsistema de acción*: son los actores que de una forma u otra implementan el programa. Por ejemplo, técnicas, dirección o colaboradoras.
- *Subsistema de reacción*: lo componen actores sociales heterogéneos, que reaccionan ante el programa, ya sea porque son los beneficiarios, porque no reciben los beneficios o porque se erigen en defensores de uno u otro grupo.

Se puede añadir a este listado un cuarto actor, que es el propio equipo de evaluación. A las evaluadoras se les confiere legitimidad para participar,

aunque con ciertas limitaciones. Cronbach advierte que «el evaluador debe buscar cuestiones procedentes de la gente del lugar; (...) se excedería en su autoridad si todas las cuestiones del estudio fueran de su invención» (Monnier, 1992: 163).

Otra fuente importante es la comunicación institucional (programación, memorias, publicaciones, producciones audiovisuales, registros, etc.). La documentación representa un determinado discurso de la institución, normalmente el de los actores con mayor poder. Otros actores sin acceso a las publicaciones, pero con capacidad de influencia sobre la práctica pueden tener otras visiones diferentes no recogidas en la documentación. Por eso el estudio de la producción documental o audiovisual normalmente es el estudio del discurso institucional.

Por último, hay perspectivas externas que pueden tenerse en cuenta e incluirse en el proceso de diseño de evaluación a través del estudio de la bibliografía especializada, de las investigaciones, de los metaanálisis, de las metaevaluaciones o de cualquier otra reflexión teórica. Es una forma de incluir un conocimiento social más amplio y de ajustar la evaluación a los criterios de la comunidad científica.

Esta propuesta de fuentes y actores no es normativa, hay otros autores y autoras que lo pueden entender de otra manera, como por ejemplo, la evaluación colaborativa de Liliana Rodríguez (2008): dicho modelo define *a priori* quién participa y cuál debe ser su rol. En otros modelos como las aproximaciones «objetivistas», o también llamadas «científicas», no se recogen las perspectivas de los diferentes actores, incluso las rechazan por entender que es una fuente de sesgo. Scriven (1991) argumenta a favor de tener el menor contacto posible con el cliente por el riesgo a ser influenciado. Aunque la crítica que se hace a estos modelos es que responden a las perspectivas de las evaluadoras enmascarándolo de supuesta objetividad.

2.3. Extracción de las necesidades informativas

El término *necesidades informativas* hace mención a los requerimientos que cada actor tiene con relación a la evaluación. El objetivo de esta fase es

detectar las necesidades de información de los actores. Se aplican técnicas de recogida de información y de análisis realizando un *primer trabajo de campo*⁶.

De esta tarea pueden emerger criterios de valor, preguntas, posibles sistemas de medición, indicadores o consejos útiles para desarrollar la evaluación, incluso predisposiciones hacia la evaluación. Para las preguntas de evaluación, Cronbach realiza una propuesta para poder seleccionarlas. Diferencia entre una fase divergente y otra convergente (Stufflebeam, 1995).

2.3.1. Fase divergente: recopilación de las necesidades informativas

Se trata de recopilar todas las posibles necesidades de información. Para extraer dichas cuestiones se puede acudir a distintas fuentes, como la documentación del programa o preguntar a los diferentes actores implicados. La labor de la evaluadora es extraer las preguntas o bien relacionadas con los criterios estandarizados, o bien con los propuestos por ellos mismos.

Tabla 1. Qué preguntas realizar

| | | |
|----------------------------|---|---|
| Modelo de criterios | Modelo preordenado | Identificación de preguntas relacionadas con los criterios de valor. Por ejemplo: ¿Qué le preocupa sobre la pertinencia del programa? ¿Qué cosas le preocupan sobre la sostenibilidad? |
| | Modelo ad hoc, sensible a los diferentes actores | ¿Qué cuestiones le parecen más importantes evaluar? ¿Qué es lo que debe valorarse de este programa? Identificación de preguntas relacionadas con los criterios de valor mencionados por ellos mismos. Por ejemplo: De la calidad que ha comentado, ¿qué cuestiones le preocupan? |

Aunque las preguntas están muy estructuradas, se suele obtener una información variada en la que se entremezclan preguntas, criterios, indicadores, consejos varios, propuestas metodológicas, impresiones, valoraciones del programa o cualquier otro aspecto relacionado con la evaluación o el programa. Toda esta información hay que depurarla y ordenarla, seleccionando preguntas, también llamadas *subcriterios de evaluación*. Para filtrar las preguntas, se pueden tener en cuenta los siguientes aspectos:

- Las preguntas deben implicar una valoración y no una mera indagación sobre algún aspecto del objeto.

- Deben estar dirigidas al programa (o a su relación con las necesidades sociales) y no a sucesos o dinámicas sociales, ya que esto sería otro objeto de estudio más propio de la investigación.
- No deben ser indicadores, sino que deben transmitir las cuestiones que preocupan a los actores implicados indistintamente de que especifique o no cómo se van a medir.

Teniendo en cuenta las salvedades hechas, la intención de esta fase es obtener un listado de preguntas cuanto más completo mejor, sin limitaciones. Los posibles sesgos no deben estar causados por la perspectiva parcial de ningún actor, o como advierte Cronbach «por la restrictiva visión del patrocinador, administrador o de quienes tengan la responsabilidad sobre el programa» (Stufflebeam, 1995: 152).

2.3.2. Fase convergente: selección de las necesidades informativas

No todas las preguntas identificadas tienen que ser contestadas por la evaluación, ya sea porque son poco importantes, porque despejan pocas dudas o porque no pueden ser asumidas por los costos que implicaría obtener los datos para responderlas. Según Cronbach se requiere de una fase previa que permita la selección de las preguntas en función de una serie de criterios, a esto lo ha llamado *fase convergente*. El autor propone dos criterios a tener en cuenta.

- *Influencia*: entre las diversas preguntas puede haber algunas más importantes para la toma de decisiones o el funcionamiento del programa, estas son las que se priorizarían en la selección.
- *Incertidumbre*: hay cuestiones sobre las que hay pocos conocimientos o estos no son confiables y, por el contrario, hay otras cuestiones que, aunque la evaluación pueda aportar conocimiento, se tiene bastante certeza y seguridad sobre cuál será la respuesta. La evaluación debería centrarse sobre las más inciertas.

Para ayudar a la selección, se puede crear una tabla multicriterio para establecer una priorización de las preguntas (ver Tabla 2).

Tabla 2. Multicriterios para priorizar las preguntas

| | Gran influencia | Poca influencia |
|--------------------|-----------------------------|-----------------------------|
| Gran incertidumbre | Preguntas muy prioritarias | Preguntas prioritarias |
| Poca incertidumbre | Solo información incidental | Solo información incidental |

El peligro de este ejercicio es que se dé más importancia a la perspectiva de unos actores sobre la de otros. La selección final de preguntas será la estructura de la evaluación, probablemente la parte más importante de todo el proceso de diseño. Todo lo demás es pensar en cómo responder a estas cuestiones identificadas.

2.4. Operacionalización

La operacionalización es un proceso deductivo que desde los criterios va estableciendo los subcriterios (preguntas), los indicadores y sus correspondientes fuentes de verificación. De forma secuencial y lógica, se va pasando por diferentes fases que desde el concepto más general y abstracto van aportando un mayor grado de concreción. Esta cadena finaliza en el diseño o búsqueda de las fuentes de extracción de información para cada uno de los indicadores.

La finalidad de la operacionalización es obtener una información, que de forma sistematizada y coherente permita establecer en cada uno de los criterios un enjuiciamiento sobre el programa.

La operacionalización se concreta en dos fases (Bustelo, 2010):

- *Trabajo vertical*: explicitación de criterios. Definición de preguntas.
- *Trabajo horizontal*: formulación de indicadores. Búsqueda de fuentes de extracción de información.

2.4.1. Trabajo vertical: definición de los criterios y preguntas

Una vez que se obtienen los criterios, ya sean preestablecidos o extraídos de los diferentes *stakeholders*, es necesario concretarlos a través de preguntas (también llamadas *subcriterios*) para que puedan ser evaluables. Los criterios reflejan un estado deseable e ideal para el programa, pero requieren al menos de un nivel más bajo y operativo que concrete ese ideal.

Aunque los criterios estén muy bien acotados y definidos, es inevitable que den cabida a múltiples significados.

Por ejemplo, la pertinencia puede estar claramente definida: «la adecuación del programa o servicio en todas sus dimensiones al contexto; entendiendo como contexto tanto las necesidades de la población, las políticas y ordenamientos jurídicos, las estructuras sociales y culturales, la teoría y el método de intervención e incluso hasta la satisfacción de los agentes implicados» (Ligero, 2010). Pero aun partiendo de esta misma definición, el criterio se expresa de múltiples formas. En la Tabla 3 se muestran dos operacionalizaciones diferentes de pertinencia basadas en la misma definición.

Tabla 3. Ejemplos de dos operacionalizaciones de pertinencia

| Programa de prevención de drogas en los colegios elaborados por técnicos locales | Programa de prevención de drogas en los colegios elaborados-importados de otro país |
|--|---|
| ¿La intervención está apoyada en una teoría sobre el cambio de los individuos? | ¿El programa está adaptado a las características socioculturales de la población? (¿se entiende?, ¿se tienen planteamientos similares ante las drogas?) |
| ¿Existe una congruencia lógica entre las actividades y los fines que se pretenden conseguir? | ¿Las evidencias o teorías asumidas están validadas para diferentes contextos y países? |
| ¿Es congruente con las necesidades manifestadas de la población? | ¿Encaja bien con el engranaje educativo local? (coordinación, horas, asignaturas) |
| ¿Es congruente con la normativa y legislación local y nacional sobre drogas? | ¿Se han apropiado del proyecto? |

Cualquiera de las dos operacionalizaciones son válidas, en los dos casos se evalúa la pertinencia, pero de forma distinta. Por lo tanto, en cada evaluación, siempre y cuando esta no sea estandarizada, se está realizando una definición particular de cada criterio.

Algunos modelos preordenados parten de preguntas ya establecidas, como es el caso del TQM-EFQM (MAP), aunque es bastante usual dejar cierta flexibilidad en la operacionalización para permitir la adaptación del modelo de evaluación al contexto. Tal es el caso de la DGPOLDE, que aconseja utilizar un modelo por criterios, pero propone que las preguntas reflejen los intereses de los actores implicados los de las evaluadoras (MAE, 2007: 61).

2.4.2. Trabajo horizontal: definición de indicadores y fuentes

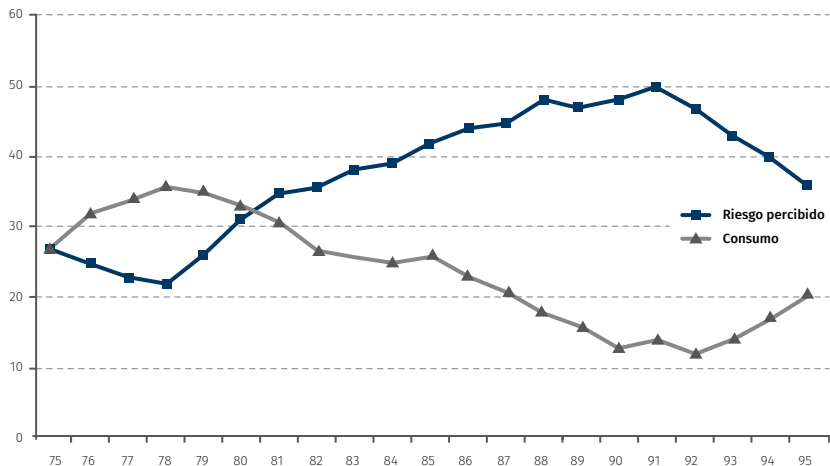
Las preguntas de evaluación son la definición del criterio, no tienen que atender al hecho de que puedan ser medidas o no. El indicador, en cambio, es el recurso que permite medir el concepto definido, por lo general a través de aproximaciones. Mientras las preguntas se tienen que encargar de representar lo más completamente posible las diversas cuestiones que definen el criterio, el indicador debe ser una variable medible, observable y registrable que aporte una información que conteste a las preguntas.

«La generalidad de los conceptos constituyen variables “latentes”, “hipotéticas”, no directamente “observables”. Por lo que su concreción precisa de la traducción del concepto teórico a indicadores, a variables empíricas (observables o manifiestas), que midan las propiedades latentes enmarcadas en el concepto» (Cea D’ Ancona, 1996: 125).

Los indicadores, por tanto, buscan acercarse lo más posible al concepto o al criterio a medir, aunque frecuentemente tienen que conformarse con reflejar la realidad de fenómenos paralelos, similares o parciales del concepto buscado aportando a la evaluación solamente una aproximación. Tal como describe Rutter (2000: 40) «los indicadores son rasgos que tienen relaciones directas con los procesos causales, pero que en sí mismos no forma parte de los mecanismos que están directamente asociados con la causalidad».

Por ejemplo, un programa de prevención escolar tiene como objetivo reducir el consumo de cannabis. A la hora de medir el consumo se plantean problemas, si se pregunta directamente a los consumidores, estos pueden falsear la respuesta para evitar ser identificados. Si, por el contrario, se buscan pruebas médicas, se necesita el consentimiento de las personas a las que se les van a practicar, y los consumidores pueden buscar estrategias para no dar su consentimiento. Aunque el consumo de cannabis sea un concepto bastante concreto, es difícil obtener la información directamente. Pero existen estudios previos como el de Bachman (2002) que muestran la alta correlación existente entre la percepción del riesgo y el consumo (ver Gráfico 1).

Gráfico 1: Correlación entre consumo y riesgo en el consumo de marihuana



Fuente: Elaboración propia a partir de los datos obtenidos en SAMHSA.

Aunque la percepción del riesgo no es el concepto en sí mismo que se quiere medir, se puede asumir que observando el riesgo, algo a lo que los jóvenes sí están dispuestos a contestar, podemos extraer conclusiones sobre el consumo. No es una relación directa, pero sí es una correlación bastante alta que permite acercarnos al concepto. La percepción de riesgo es un instrumento, una variable medible que nos aproxima al concepto que se quiere evaluar, es decir, un indicador.

Los indicadores deben ser redactados como condiciones lógicas que si se cumplen, contestan a la pregunta de evaluación. Se fija un estado, un punto o una condición a partir del cual se puede contestar afirmativa o negativamente a la pregunta formulada. En el caso del consumo de cannabis, la redacción del indicador podría ser el que se indica en la tabla siguiente (ver Tabla 4).

Tabla 4. Ejemplo de pregunta de evaluación e indicador

| Pregunta de evaluación | Indicador |
|--|---|
| Disminución del consumo de marihuana entre los escolares | Incremento con relación al semestre pasado de la percepción del riesgo en el consumo de marihuana/hachís. |

La condición, que en el ejemplo citado es «incremento con relación al semestre pasado», es el estándar. Un estándar es «una cantidad, un nivel o una manifestación de un criterio determinado que indica la diferencia entre dos niveles de mérito distintos» (Stake, 2006: 111). Los estándares pueden ser dicotómicos o establecer gradientes, ser flexibles o estrictos, relativos o absolutos, cualitativos o cuantitativos, pero de alguna forma tendrán que explicitar en qué punto o momento la pregunta de evaluación es contestada.

2.5. Consideraciones sobre la evaluación por criterios

La secuencia lógica de una evaluación por criterios, la que se ha desarrollado en este capítulo, es la que propugna Fournier (Stake, 2006: 242).

- Establecimiento de criterios sobre el mérito del programa: ¿en qué dimensiones debe funcionar bien el evaluando?
- Definición de estándares: ¿cuál es el funcionamiento mínimamente bueno que debe alcanzar el evaluando?
- Medición del rendimiento y comparación con los estándares: ¿hasta qué punto funcionó bien el evaluando?
- Síntesis e integración de los datos en un juicio o valoración del mérito de la valía: ¿cuál es el mérito o la valía del evaluando?

Según esta forma de trabajar la evaluación, el producto final que se obtiene es una serie de juicios en las áreas definidas previamente (criterios). Es decir, si la información obtenida por los indicadores es viable y fiable, si se logra el nivel definido por los estándares, entonces las preguntas serán contestadas afirmativamente y, por consiguiente, el juicio sobre el programa será positivo. El programa será eficaz, eficiente, pertinente, viable, etcétera.

Hay dos aspectos de esta secuenciación para el enjuiciamiento que es interesante remarcar. La primera es que el juicio se establece en los estándares de los indicadores, son realmente los criterios últimos de valoración. Los grandes criterios son solo áreas de análisis dentro de las cuales los estándares trabajan. Por lo tanto, la definición de los indicadores

no es un mero trabajo técnico, sino que son los que determinan la valoración final del programa.

La segunda cuestión es que esta dinámica valorativa conduce a un enjuiciamiento sumativo sobre el programa (según la clasificación de Scriven). Por ejemplo, la evaluación de un proyecto de educación profesional básica apoyada en la estrategia de la alternancia (períodos en el centro educativo y períodos en el medio socio-profesional) en una zona rural andina⁷ fijó una serie de criterios: impacto, eficacia, eficiencia, cobertura, sostenibilidad, apropiación y calidad. Se establecieron unas preguntas y unos indicadores con sus respectivos estándares. Tras la realización del trabajo de campo y el análisis de los datos se obtuvieron las siguientes valoraciones:

- Detectaron varios impactos positivos (los cuales son relatados aunque no son concluyentes).
- Valoraron que la eficacia, la eficiencia, la cobertura y la calidad eran altas.
- La apropiación del proyecto se consideró media.
- La valoración de la sostenibilidad queda en suspenso por falta de información en el momento de hacer la evaluación.

Además de estos juicios, el informe de evaluación describe el procedimiento realizado, cómo se han obtenido los datos, la interpretación de estos para llegar a los juicios y propone recomendaciones y lecciones aprendidas. El informe puede dar para más lecturas que sin lugar a duda resultarán interesantes, pero realmente el producto final de este modelo es la valoración de los siete criterios expuesta en los cuatro guiones anteriores.

Lo que produce este tipo de evaluación, además de las recomendaciones, es enjuiciamiento. Algo que guarda una estrecha relación con el rendimiento de cuentas público, también llamado *responsabilización*.

2.6. Críticas a la evaluación orientada por criterios y estándares

Críticas referidas a los modelos de criterios preordenados (o estandarizados):

1. *Se considera que es una evaluación parcial*: como se ha visto anteriormente, existe un gran consenso en la literatura sobre que los diferentes grupos de interés implicados en la evaluación no tienen por qué compartir los mismos intereses, necesidades informativas o criterios sobre la calidad de un programa. El propio concepto de calidad varía dentro y entre los diferentes grupos de *stakeholders* (Greene, 1999; Stake, 1967). Por tanto, los modelos preordenados y los estándares no suelen reflejar de forma homogénea los intereses evaluativos o lo que consideran bueno los diferentes actores. Asumir una forma estándar para valorar una intervención es responder a la visión de un determinado punto de vista que puede ser que no coincida con la de los técnicos, los beneficiarios u otros actores.

Los enfoques integrados propuestos por Rossi y Chen (1990) intentan superar esta primera crítica sugiriendo una negociación entre los actores clave para obtener un diseño de evaluación de consenso y que satisfaga las diversas demandas. Pero Stake (Shadishy Leviton, 1991) advierte que la evaluadora no debe crear un consenso inexistente. Aunque se lograra un acuerdo, fallaría en su aproximación a las realidades diversas de una sociedad (Stake cit. Greene, 1999: 164). La aplicación de estándares universales incluso puede entenderse como la plasmación de una determinada forma de autoridad (Abma cit. Greene, 1999).

Los modelos preordenados son métodos parciales con una utilidad limitada para el conjunto de actores, que no tienen en cuenta las diferentes prioridades e imponen una determinada visión de la calidad y el valor de la intervención.

2. *Evaluación sesgada. Pretensión de objetividad*: muy relacionado con el punto anterior es que los modelos preordenados, en la medida en que son propuestas explícitas, generalizadas y asumidas por diferentes actores, corren el riesgo de ser tratadas como «objetivas» e independientes de los propios actores que las formulan. Por ejemplo, los cinco criterios del CAD representan un determinado sistema de valores, el de la OCDE, organización de los Estados

más industrializados del mundo y principales donantes de la cooperación internacional. Es lógico pensar que para ellos el criterio de sostenibilidad es muy importante, ya que se fija en que la ayuda sea coyuntural y no estructural, es decir, no permanente.

Pero gran parte de las políticas (muchas de ellas objeto de cooperación internacional: asistencia sanitaria, asistencia social, mantenimiento del orden, seguridad y seguridad jurídica, entre otras) no son por definición sostenibles ni pretenden serlo. Sus objetivos procuran satisfacer necesidades impostergables o impedir (o minimizar) perjuicios derivados de circunstancias adversas. En definitiva, mantener bajo control la desorganización que sobrevendría si no se lograran. Son intervenciones indispensables, necesarias pero no sostenibles desde un punto de vista economicista (Hintze, 2005: 147). El autor declara que «las medidas simples y puntuales que cada tanto producen efectividades sostenibles como por arte de magia, parecen ser precisamente eso: arte de magia y como tales no parecen demasiado frecuentes en el mundo de la gestión de lo público».

La inclusión del criterio de sostenibilidad en el análisis de la acción pública implica una particular visión de cómo deben ser y cómo deben funcionar las políticas y proyectos financiados, lo que quizá no encaje con la visión de los poderes públicos locales o de los beneficiarios. La perspectiva del CAD es legítima pero interesada, responde a su propia visión e intereses.

El peligro no está en que se propongan determinados criterios de evaluación, sino en que se entiendan estos como métodos objetivos y permanezcan en la sombra las intencionalidades que los motivan, algo que Abma ha definido como «autoría enmascarada» (Greene, 1999: 169).

Para evitar este sesgo metodológico, existen propuestas, como por ejemplo la de Scriven, que invitan a alejarse de los clientes, de su encargo y de la forma en que hay que hacer evaluación. Pero esto no hace más que sustituir los criterios del cliente por los de la evaluadora, algo que Macdonald (Stake, 2006) ha denominado «evaluadora autocrática».

Si volvemos la mirada hacia los autores apoyados en paradigmas constructivistas, encontramos que el valor de una intervención está depositado en el que mira y no reposa en la sustancia del objeto evaluado (Greene, 1999: 3). Incluso Scriven, uno de los mayores buscadores de objetividad en el campo de la evaluación, entiende que esta se alcanza a través de un acuerdo intersubjetivo (Chen, 1990: 62).

El «pensamiento negociado» necesario para llegar a un valor compartido sobre la calidad y el valor requiere tiempo (Stake, 2006). Cuando los criterios están claramente definidos y estructurados y no permiten una reinterpretación o transformación de los puntos críticos de la calidad del programa por otros actores implicados, se está ante una imposición del valor en ocasiones camuflada de objetividad.

3. *Los modelos de criterios tienden a centrarse casi exclusivamente en la dimensión de resultados:* en el modelo CAD, el impacto y eficacia miden exclusivamente el cumplimiento de la jerarquía de los objetivos. La eficiencia analiza la relación de dichos resultados con los «*inputs*» del programa. La sostenibilidad quiere valorar el mantenimiento de la propia intervención y de sus resultados a largo plazo. Y, por último, la pertinencia según la entiende el CAD se centra principalmente en la coherencia de los objetivos del proyecto con los objetivos de los donantes y del gobierno. Tal como se puede ver, en los cinco criterios, la dimensión clave son los objetivos/resultados sobre la cual pivotan los juicios.

A las evaluaciones de resultados se las ha denominado de «caja negra» (Chen, 1990; Weiss, 1998). Se caracterizan por mirar solo a los resultados o como mucho a los *inputs*, sin preocuparse por el proceso intermedio de transformación.

La evaluación de resultados no se fija en la implementación ni en aspectos estructurales, por lo tanto «falla en identificar las deficiencias para futuras mejoras y desarrollos del programa» (Chen, 1990: 18). Debido a estas dificultades Stufflebeam (2001) las denomina cuasievaluaciones, ya que emiten una valoración de un todo, el programa, valorando solo una parte, los resultados.

4. *Problemas con los objetivos*: el término *objetivo* tiene diferentes acepciones que lo convierten en un concepto engañoso. Por un lado, la palabra se refiere a la meta, fin, misión, en definitiva a la «diana» que tiene que alcanzar el programa. Pero *objetivo* también significa algo ajeno a las percepciones, no expuesto a las subjetividades, es decir, sin sesgo. Los objetivos son propuestos por actores que voluntaria o involuntariamente representan determinadas posiciones. Si se pide a los diferentes actores de una organización que definan los objetivos, no los que están en los documentos, sino los que realmente quieren alcanzar, es probable que muchos de estos no coincidan. Ralph Tyler, en su evaluación «Eight Year Study» sobre 30 *high schools*, encontró que en el primer año solo dos o tres centros estaban trabajando de acuerdo a lo descrito en el programa (Weiss, 1998: 9).

En la literatura se encuentran muchas otras referencias que advierten sobre la distancia entre los objetivos «oficiales» y los que al final se aplican. Stake habla de «objetivos declarados», aludiendo a aquellos que están en el papel que no tienen por qué coincidir con los que realmente trabaja la institución. Los objetivos, entiende Monnier, son la resultante de un proceso de negociación. Los objetivos declarados, los que están en la documentación o matrices de programación, suelen representar la visión de los actores con más poder. Incluso se llega a entender la política «como una declaración de intenciones por parte de actores que desean convencer o controlar a otros actores» (Deleau cit. Monnier, 1992: 105).

En ocasiones, se superan las tensiones entre los diferentes actores implicados en una intervención, llegando a acuerdos sobre los objetivos superiores (general y específico) con formulaciones generales trasladando el disenso a los objetivos operativos (resultados esperados).

Estos condicionantes hacen que en un proceso de evaluación «los objetivos reales no se obtienen en documentos o en conversaciones, mejor ver a qué dedican su tiempo los actores implicados» (Weiss, 1998: 55), ya que se puede llegar al extremo de que la institución trabaje realmente en pro de unas metas que disten mucho de parecerse a las que tiene escritas; algo bastante alejado del concepto de metas insesgadas.

Críticas a todos los modelos de criterios y estándares:

5. *Problemas de los estándares*: hay que tener en cuenta que en el caso de los criterios de eficacia e impacto, los estándares son los objetivos declarados.
 - *Reducción del valor del programa a un punto final*. El punto final o estándar simplifica el juicio, pero también reduce a una sola medida la complejidad de una intervención. «Podemos medir la productividad de una persona que envasa mantequilla, pero de nada nos sirve si para hacerlo se lame el pulgar antes de coger el papel de envasar» (Messick cit. Stake, 2006: 248). Es un proceso de reducción y de pérdida de información sobre la variada y rica realidad de la intervención y rara vez mejora el conocimiento sobre lo que se hace.
 - *Alcanzar el estándar*. Si no se alcanza el estándar fijado, no se puede emitir un juicio positivo sobre la intervención aunque el trabajo realizado haya sido encomiable. Si, por ejemplo, en un determinado indicador se fija un cumplimiento de 100 % y se obtiene 98 %, el juicio que se emita sobre ese aspecto debe ser negativo aunque se haya logrado un 98 %. En este caso, dos puntos pesan más que los 98 restantes.
 - *Estándares desajustados*: si los estándares declarados son ambiciosos, el programa no será eficaz aunque el dispositivo sea excelente. Por el contrario, si los estándares son muy humildes, siempre habrá una excelente versión de lo que ha ocurrido aunque el programa esté en mínimos e infrautilizado.
 - *Legitimidad de quién fija los estándares*: teniendo en cuenta que sobre los estándares recae el peso del enjuiciamiento final, ¿quién o quiénes están legitimados para fijar los estándares? En ocasiones, se encubre como una mera decisión técnica lo que es de calado estratégico o político. Stein encuentra que frecuentemente los líderes políticos encubren el debate de cuestiones importantes sobre el valor del programa en supuestamente neutrales decisiones técnicas acerca de los indicadores y las medidas (House, 2006: 126).

6. *Visión túnel*: al fijarse la evaluación en unos determinados criterios, pueden pasar desapercibidos otros elementos importantes de la intervención o sus resultados. Este efecto se ha definido como «visión túnel» (Leeuw y Furubo, 2008). La crítica no se ciñe solo a una pérdida de la riqueza de evaluar, sino que evaluar solo lo que está escrito puede ser una forma elegante de controlar a la evaluadora. Un centro de menores infractores puede tener definidos claramente los aspectos que se pretenden alcanzar relacionados con la eficacia, el impacto o la pertinencia conformando una evaluación completa. Pero puede que no hayan tenido en cuenta y, por lo tanto, no habrán indagado nada sobre casos de palizas y malos tratos. Scriven procura evitar el sesgo que producen tanto el cliente como el encargo, con su modelo de evaluación libre de objetivos, centrándose en los cambios en la realidad sin que esto tenga que estar estructurado por la declaración de objetivos.
7. *Enjuiciamiento, no comprensión*: probablemente la crítica más importante al método de evaluación por criterios es que no conduce hacia la comprensión de lo que ocurre en el programa, sino hacia una valoración sumativa.

El establecimiento de unos criterios de valor, con sus consiguientes estándares (ya sean más o menos estrictos o laxos), genera una cadena lógica que concluye en un juicio sumativo en cada una de las áreas definidas. Como se ha argumentado en el apartado «Consideraciones sobre la evaluación por criterios», se obtienen unas valoraciones y el programa se adjetiva como eficaz o ineficaz, pertinente o impertinente, etc. Obtener valoraciones de un programa tiene interés social y político; en sí mismo justifica una evaluación y contribuye directamente a la responsabilización o *accountability* de la acción pública.

La limitación radica en que esta estrategia no provee de explicaciones de cómo han ocurrido las cosas. Poniendo un ejemplo sencillo, se puede valorar el desempeño de una estudiante y concluir que no ha sido eficaz y ha suspendido el curso. Podemos agregarle más criterios y concluir que es pertinente su plan de estudios, porque sus objetivos son coherentes con el centro académico y el método escogido está

avalado por la teoría. Pero no nos permite saber más, no sabemos lo que está pasando con el desempeño, quizá la estudiante no sepa aplicar bien las técnicas de estudio, no esté motivada, sea miope, tenga una agenda muy cargada, sufra ansiedad o los profesores no sean buenos. Todas estas dimensiones, factores causales y aspectos relevantes han quedado reducidos a una valoración dual y sumativa.

Se puede buscar una interpretación más pormenorizada e intentar entender qué hay debajo del juicio, nadie lo va a impedir, pero la mecánica establecida en el método de criterios no está dirigida a ello, sino a la emisión de juicios. La finalidad comprensiva de la evaluación queda relegada frente a la valoración y el enjuiciamiento y por tanto está más dirigida al rendimiento de cuentas.

El énfasis en el rendimiento de cuentas público (*accountability*) habla también de unas determinadas inquietudes sociales frente a otras. Hace treinta años, Lee Cronbach observó que la demanda de *accountability* es un signo de patología social en el sistema político. Jennifer Greene apunta que lo considera una demanda regresiva e irrelevante para su trabajo como evaluadora (Greene, 1999: 1).

Lo que remarcan estos autores es que enjuiciar sin comprender obedece a una necesidad de catalogar, enjuiciar o «colocar programas en esquemas» más que a una motivación constructiva de mejora. La mecánica del modelo de criterios tiende a eso, a valorar sin la necesidad de pasar por la comprensión.

3. Evaluación de la teoría del programa

De forma resumida se puede definir la evaluación de la teoría del programa como un proceso que, en primer lugar, identifica cómo funciona la intervención y qué resultados pretende obtener construyendo un modelo teórico del programa y, en segundo lugar, estructura todo el proceso de indagación a través de dicho modelo teórico. No solamente se va a evaluar si se han logrado los objetivos, sino que también se podrán identificar los mecanismos causales que han conducido a ellos.

En este capítulo se pretende describir cuáles son las fases y las reflexiones teóricas principales para el diseño y la realización de una evaluación bajo este prisma. Se incorporan al final unos epígrafes sobre ventajas y debilidades.

Primeras referencias teóricas

La evaluación de la teoría del programa tiene un amplio bagaje teórico. La idea básica ya se puede encontrar en los años 60 en Schuman (cit. en Rogers, Petrosino, Huebner y Hacsí, 2000: 6) cuando sugiere que la evaluación podría dirigirse hacia la cadena de objetivos de la intervención. La misma idea se repite en Bennet en el año 1975, quien utiliza el concepto de *jerarquía de resultados* (citado en Rogers, Petrosino, Huebner y Hacsí, 2000). Progresivamente se va extendiendo la idea de que para evaluar es necesario el desarrollo de un marco teórico previo del programa.

En 1987 Bickman edita una publicación titulada *Using Program Theory in Evaluation*, en la que define la teoría del programa como la construcción de un modelo razonable y sensible de cómo se supone que funciona la intervención. En la misma publicación, Wholey señala que el propósito de la teoría del programa es identificar los recursos del programa, las actividades y los resultados previstos y especificar las cadenas de asunciones causales vinculando recursos, actividades, resultados intermedios y objetivos finales.

Desde 1972 Carol Weiss invita a identificar posibles modelos causales de los programas. Ella lo denomina teoría del cambio y lo describe como «la cadena de asunciones que explican cómo las actividades conducen paso por paso a los resultados esperados» (1998: 2).

Schon, en 1997, utiliza el término *teoría de la acción*, muy parecido al usado por Carol Weiss y los autores Lenne, Cleland y Funnell, que la denominan «lógica del programa» (Rogers *et al.*, 2000).

«Evaluación dirigida por la teoría» (*Theory-driven Evaluation*) es el título dado por Chen para una de las monografías de referencia en este enfoque. Lo define como la especificación de «lo que debe ser hecho para alcanzar la meta deseada. Que otros importantes impactos pueden ser anticipados y como esos objetivos e impactos podrían ser generados» (Chen, 1990: 40).

Davidson (2000) prefiere denominarla «evaluación basada en la teoría», para evitar transmitir la idea de que la evaluación pueda estar condicionada *a priori* por una determinada teoría.

La palabra *teoría* despierta también otro tipo de recelos. Rossi (Prefacio escrito por Rossi en Chen 1990: 9) la considera demasiado pretenciosa para describir los mecanismos internos que operan en un programa. La alternativa a teoría es el modelo lógico, en ambos casos se refieren a la cadena de asunciones causales que hay detrás de cada intervención. En este texto se utiliza de forma indistinta teoría o modelo lógico del programa.

Los postulados de la teoría del programa están incorporados en diversas perspectivas teóricas. Bob Stake se declara abiertamente cercano a esta forma de proceder. Rossi ve necesario la identificación de las relaciones lógicas en las intervenciones e incluso el experimentalista Tomas Cook se declara fan de la evaluación orientada por la teoría (2000: 32).

Fases para el diseño de una evaluación orientada por la teoría

Tal y como se ha definido anteriormente, la evaluación orientada por la teoría consiste en explicitar la teoría o modelo lógico del programa y las relaciones causales entre todos sus componentes y orientar la evaluación según dicho modelo.

Los pasos usuales en este tipo de evaluación son plasmar en un gráfico las relaciones causales que van desde los *inputs* a través de los procesos hasta provocar los resultados (Stufflemban, 2001). Posteriormente, se identifican las dimensiones clave que deben ser medidas. Con la información que se obtiene se pueden poner a prueba las relaciones causales y extraer conclusiones sobre el modelo propuesto. Es decir, se trata de indagar sobre si los éxitos o fracasos del programa son debidos a las diferentes cadenas de trabajos y si los medios son adecuados para la implementación. Se proponen una serie de fases:

Fases de la evaluación orientada por criterios:

- 3.1. Construcción, reconstrucción o identificación de la teoría del programa o modelo lógico.
 - 3.1.1. Resultados.
 - 3.1.2. Dimensión de procesos o mecanismos de implementación.
 - 3.1.3. Dimensión estructural: los recursos, la organización y otros elementos estructurales.
 - 3.1.4. La perspectiva sistémica.
- 3.2. Múltiples perspectivas sobre las teorías que subyacen a una intervención.
 - 3.2.1. Fuentes para obtener las teorías que subyacen a una intervención.
- 3.3. Modelo lógico con perspectiva sistémica.
- 3.4. Graficar la teoría.
- 3.5. Determinación de los aspectos a evaluar.
- 3.6. Métodos y técnicas.
- 3.7. Análisis.

3.1. Construcción, reconstrucción o identificación de la teoría del programa o modelo lógico

En esta fase el objetivo es identificar, reconstruir o elaborar directamente el marco teórico que establece relaciones causales entre las diferentes dimensiones de una intervención. Se van a destacar tres dimensiones clave: resultados, procesos y elementos estructurales.

3.1.1. Resultados

Probablemente, el aspecto más requerido en evaluación son los resultados. Los resultados son los cambios en la realidad producidos o atribuibles a la intervención. Hay tres formas principales de acotar los resultados:

- a. Utilizar resultados estandarizados.
- b. Fijarse en los resultados definidos por los objetivos del programa.
- c. Observar todos los posibles cambios dados en la realidad.

En el caso «a», los resultados que se evalúan son los mismos para diferentes intervenciones. Un buen ejemplo es el informe sobre el sistema educativo PISA. Se comparan los rendimientos académicos de los estudiantes de diferentes países ante pruebas iguales. En el sector de las ONG, el informe de la Fundación Lealtad⁸ se apoya también en una lógica estandarizada. Todas las ONG son analizadas bajo los mismos baremos indistintamente de las peculiaridades de cada una.

El caso «b» es el modelo más extendido desde que Ralph Tyler en 1942 propuso la evaluación de los objetivos como medida del éxito del programa (Stufflebeam and Shinkfield, 1995: 93). Está presente en la propuesta de criterios del CAD (eficacia e impacto). Los objetivos son soluciones proyectadas, metas fijadas por el programa de antemano. Si los objetivos son un deseo, una intención, los resultados son los cambios constatados provocados por la intervención. Como se espera que los resultados sean la proyección de los objetivos, en ocasiones los términos *objetivos* y *resultados* se utilizan de forma indistinta, al menos en determinados sistemas de programación⁹, gestión o evaluación. Pero estrictamente los objetivos son meras intenciones y los resultados realidades, en la práctica no tienen por qué coincidir.

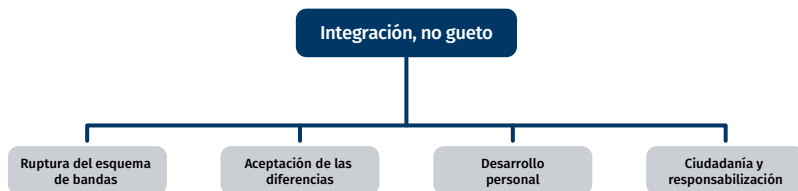
Lo primero, entonces, será identificar o programar los resultados pretendidos para poder evaluarlos. Los métodos de programación por objetivos¹⁰ funcionan básicamente proponiendo un objetivo específico y otros intermedios. Estos últimos son un medio necesario e indispensable para alcanzar el objetivo específico. La jerarquía de objetivos establece con claridad unas relaciones causales, si se logra el nivel inferior se tendrá que lograr el superior.

El soporte de programación suele ser un diagrama o una matriz donde se especifican los diferentes niveles de objetivos, desde el más genérico hasta el más concreto. Por ejemplo, en el caso de un programa que tiene como objetivo específico «la integración de los jóvenes para que no se conforme un gueto en un determinado barrio de una gran ciudad», establece unos objetivos intermedios que son:

- 1) Los jóvenes van con gente que no es solo de su misma nacionalidad o etnia.

- 2) Aprenden a aceptar las diferencias.
- 3) Incrementan su desarrollo personal.
- 4) Se responsabilizan con el entorno y la ciudadanía.

Gráfico 2: Objetivos del Programa Tiempo Joven



Fuente: Elaboración propia.

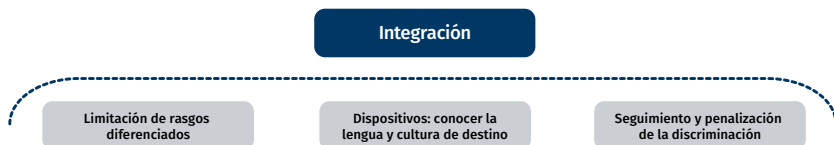
Todos los objetivos que aparecen en el diagrama (ver Gráfico 2) son cambios en la población o en el entorno. Todos son resultados, aunque los del nivel inferior son meros instrumentos para lograr el superior. En esta intervención y en este proyecto el desarrollo personal, por ejemplo, solamente importa por su cualidad de ser medio para el fin, pero en sí mismo carece de finalidad. Cada nivel inferior tiene que ser la condición suficiente y necesaria para que se cumpla el objetivo del nivel superior. De esta forma, se construye una cadena lógica, la modificación de una serie de «variables intermedias» (objetivos intermedios o resultados esperados) deben provocar cambios en el nivel superior (objetivo específico). El Enfoque de Marco Lógico es un ejemplo de esta lógica de programación.

El caso «c» trata de liberarse de las influencias y limitaciones de gerentes o financiadores u otros grupos de poder. Scriven propone observar todos los posibles cambios dados en la realidad. Se buscan directamente los resultados obtenidos por la intervención, sin conocer previamente los objetivos. Esta propuesta también conlleva críticas ya que la identificación de resultados es susceptible de estar sesgada por la propia mirada de quien los busca, en este caso de la evaluadora.

Sea cual sea la forma de acotar o identificar los objetivos se está haciendo un ejercicio de plasmación teórica, se define una relación causal entre diferentes niveles de objetivos. En el gráfico anterior (ver Gráfico 2), se

proponía una forma de conseguir la integración, pero puede haber otras formas de lograrlo. El modelo de la asimilación cree conseguirlo de esta manera (Giménez, 2007) (ver Gráfico 3).

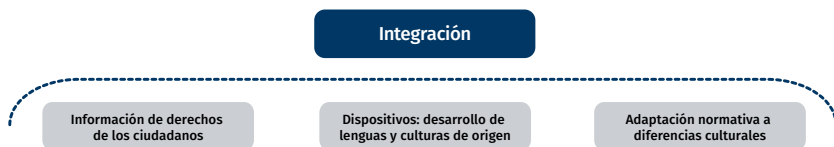
Gráfico 3. Asimilación



Fuente: Elaboración propia.

El multiculturalismo es otra forma de buscar la integración, lo cree conseguir a través de informar sobre los derechos, desarrollar las lenguas y culturas de origen y adaptación de la normativa a las diferencias culturales (Giménez, 2007) (ver Gráfico 4).

Gráfico 4: Multiculturalismo



Fuente: Elaboración propia.

Es extraño encontrar un área de intervención humana donde no se pueda abordar un mismo problema desde diferentes estrategias. Ocurre con la promoción de emprendedores, con el incremento de la renta, con la mejora de la calidad de vida, con la mejora de las competencias técnicas y educativas, en atención sanitaria, en la reducción del crimen, en la inserción de excluidos, etc. En prevención de drogodependencias, por ejemplo, Petratis, Flay y Miller (Becoña, 1999: 76) encuentran 14 formas diferentes de prevenir el consumo (ver Tabla 5).

Tabla 5. Las 14 teorías que consideran Petratis, Flay y Miller (1995) para explicar la experimentación con el uso de sustancias

| | |
|---|--|
| Teorías cognitivo-afectivas | 1. Teoría de la acción razonada. 2. Teoría de la conducta planificada. |
| Teorías de aprendizaje social | 3. Teoría del aprendizaje social. 4. Teoría del aprendizaje social/cognitivo-social. |
| Teorías del apego social | 5. Teoría del control social. 6. Modelo de desarrollo social. |
| Teorías donde las características interpersonales juegan un papel esencial | 7. Modelo de ecología social. 8. Teoría del autodesprecio. 9. Modelo de varias etapas de aprendizaje social. 10. Teoría de la interacción familiar. |
| Teorías que integran constructos cognitivo-afectivos, de aprendizaje, compromiso y apego, en intrapersonales | 11. Teoría de la conducta/problema. 12. Teoría de <i>cluster</i> de iguales. 13. Modelo de vulnerabilidad de Sher. 14. Modelo de Dominio. |

De cada una de estas propuestas teóricas se podría extraer al menos un programa de prevención con el mismo objetivo final y con diferentes objetivos intermedios. Por tanto, cuando se está frente un programa, se está ante una determinada elección teórica entre una variedad de posibilidades. A esto es a lo que Carol Weiss llama *teoría del programa* (Weiss, 1998: 55), la asunción teórica por la que opta un programa para lograr el objetivo final de una determinada intervención.

Limitaciones de la dimensión objetivos/resultados

La perspectiva de resultados presenta algunas limitaciones. No cabe duda de que conocer el grado de cumplimiento de los objetivos tiene un gran valor. Así, los que tienen que tomar las decisiones acerca del proyecto pueden saber si este ha tenido éxito o no. Pero aparte de saber si el proyecto está o no está funcionando, no se puede saber nada más.

En el caso de que se obtengan malos resultados, se desconoce si estos son producidos por una mala praxis, por una inadecuada formación del personal, por unos recursos insuficientes o por una mala gestión, entre otros posibles motivos. Al evaluar resultados se obtiene información sobre los resultados sin preocuparse por el proceso de transformación interno que los provoca. Este tipo de evaluación se denomina de «caja negra» (Chen, 1990: 18).

Con la evaluación de resultados no se pueden tomar decisiones de mejora porque no se sabe cuál es el problema. Esta es la crítica que hacen Weiss y Rien a este tipo de evaluación (Ballart, 1992: 329). Stufflebleman y Shinkfield (1995) definen las evaluaciones de objetivos como cuasievaluaciones, ya que no llegan a comprender la totalidad de la evaluanda y emiten un juicio sobre el todo analizando solo una parte.

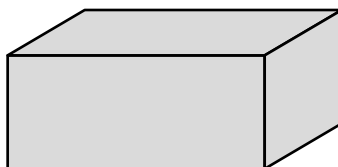
Para superar estas críticas, las evaluaciones tienen que ampliar la perspectiva con la que abordan el objeto y empezar a mirar en el interior de la caja negra. Una de las propuestas para iluminar la caja es el análisis de los procesos.

3.1.2. Dimensión de procesos o mecanismos de implementación

Le dijo la cigarra a la hormiga: «amiga mía, todavía no entiendo cómo hago para pasar todos los inviernos hambre y frío mientras tú siempre tienes comida y calor».

En la caja negra (Gráfico 5), está todo lo que se hace y ocurre en el programa; son acciones, decisiones, recursos, personas e interacciones. Se espera que la combinación de todos estos factores produzca los resultados previstos. Realmente los programas son esto (recursos, personas, interacciones, acciones, etc.), ya que los resultados son meramente las salidas, que incluso se puede llegar a conceptuar como algo externo porque es lo que «sale» de los límites de la intervención.

Gráfico 5. Caja negra



Fuente: Elaboración propia.

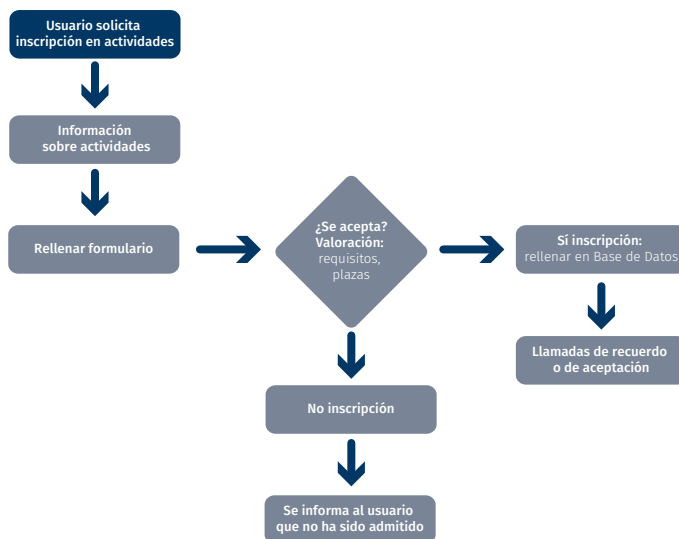
Si la perspectiva de resultados se fija en estas «salidas», la de implementación se centra en los mecanismos internos de producción, en cómo suceden las cosas. Una de las formas más comunes para trabajar con todo lo que tiene que ver con la implementación es a través del análisis de los procesos.

Un proceso es una secuencia de actividades que genera un valor para los beneficiarios respecto a una determinada situación inicial. A través de una cadena de trabajo se produce algo, ya sea un producto, un servicio, una orientación o un cambio en la población que antes no existía. Por ejemplo, una mesa ha sido generada por una secuencia de actividades que han transformado las materias primas (madera e ideas) en el producto final (mesa); lo que aporta un nuevo valor al usuario es poder escribir sobre ella o apoyar una taza de té. Si una secuencia de actividades no aporta nada, no se está ante un proceso, sino ante un mero compendio de acciones.

El término *proceso* lleva implícita la idea de sucesión, secuencia, eje temporal que organiza una serie de eventos hasta la consecución de un fin¹¹. En la perspectiva de resultados no se requiere específicamente de una cronología, se programa lógicamente pensando en qué es lo que se necesita tener o conseguir para lograr una meta superior. En cambio, con la perspectiva de procesos tiene una importancia crucial el orden en el que suceden las cosas. Hay que especificar cómo aparecen las acciones y cómo se vinculan entre sí.

Se pueden graficar los procesos con un esquema denominado *diagrama de flujo* o *flujograma*, donde las diferentes acciones se visualizan como un flujo continuado. En el Gráfico 6 se muestra un esquema sintético de uno de los procesos de un proyecto de formación (proceso de inscripción en las actividades).

Gráfico 6. Proceso: inscripción en actividades de formación



Fuente: Elaboración propia.

Con los procesos, igual que con los resultados, ocurre que no hay una única forma de hacer las cosas. Habrá unos programas que planteen una forma de llegar a los fines y puede ser que otros lleguen a la misma meta de forma totalmente diferente. La existencia de diferentes opciones resalta la idea de que se está ante una determinada elección teórica; en este caso no de logros finales, sino de métodos de actuación.

La elección escogida por el programa es lo que Weiss denomina teoría de la implementación (Weiss, 1998: 58), que se puede definir como la cadena de asunciones teóricas que vinculan la acción desarrollada por el programa o el servicio con el logro final de los resultados.

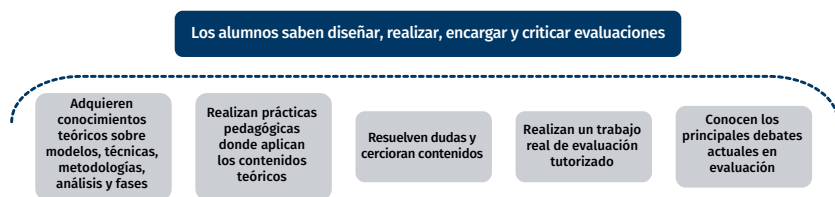
Programación por objetivos y por procesos

La principal diferencia entre resultados e implementación es que el primero describe cambios en la población o en el entorno, mientras que el segundo describe las cosas que hace el programa para conseguirlos. Por ejemplo, en un programa formativo, el objetivo es que una población adquiera

unos conocimientos. La implementación es todo aquello que se hace para lograrlo, como puede ser dar clase. Dar clase no es lo mismo que adquirir formación, se pretende que una cause la otra, pero son aspectos distintos. Esta es la diferencia básica entre implementación y resultados, lo que el programa hace y lo que el programa consigue.

Veamos otro ejemplo: en un magíster de Evaluación de Programas, los resultados son cambios en la población en sus dos niveles (ver Gráfico 7).

Gráfico 7. Ejemplo de objetivos de un magíster en Evaluación de Programas y Políticas Públicas



Fuente: Elaboración propia.

Desde la perspectiva de resultados, el gráfico en sí mismo es completo (se podría añadir un nivel más que especificara las actividades e incluso un cronograma). Pero esta forma de programar no está especialmente pensada para recoger todas las secuencias de actividades necesarias para el funcionamiento del magíster.

En la siguiente tabla, la 6, se muestra un «mapa de procesos» en el que se enuncian los principales procesos de la organización.

Tabla 6. Ejemplo de mapa de procesos. (Mapa de procesos del magíster en Evaluación de Programas y Políticas Públicas; elaborado a partir de la Memoria de Grado de Irene Rosales)

| | |
|------------------------------------|--|
| Procesos estratégicos | Definición posicionamiento estratégico. Definición operativa del programa del magíster. Organización básica. |
| Procesos clave u operativos | 1. Difusión del magíster. 2. Proceso Educativo a) Selección de alumnos/as. b) Programa de sesiones. c) Docencia. d) Trabajo intersesiones (trabajos de alumnos/as). e) Memoria de Grado. f) Prácticas en empresas. g) Trato con el alumnado. |
| Procesos de soporte o apoyo | - Selección de coordinación administrativa. - Secretaría: preinscripción y matrícula. - Administración: facturación. - Gerencia: economía y presupuestos. - Gestión de aprovisionamiento de viajes de profesores/as. - Gestión académica: calificación del alumnado (actas); control de asistencia; evaluaciones de las sesiones. - Gestión de aulas. - Gestión de convenios de prácticas del alumnado. - Gestión de recursos web. |

Aunque los procesos se quisieran incluir como listados de actividades dentro de la matriz de objetivos, esto no sería posible porque muchos de ellos son transversales y atañen a varios objetivos a la vez. Además, tampoco permitirían incorporar las divisiones de flujos en dos o más caminos tan necesarias en los diagramas de procesos.

Tanto el enfoque de resultados como el de procesos aisladamente no cubren toda la realidad de una intervención, son complementarios. La programación conjunta de objetivos y procesos incrementa la congruencia del modelo teórico. La vinculación entre procesos y resultados es probablemente una de las contribuciones más importantes a la teoría de la evaluación (Greene, 1999; Shadish y Leviton, 1991), ya que busca la articulación causal entre lo que hace el programa y lo que consigue.

Con la especificación de la lógica de la implementación, los resultados ya no brotan de una caja negra, sino que son el producto de diferentes secuencias de actividades que estarán plasmadas en los diagramas de flujos.

La combinación de la teoría de la implementación y la teoría del programa conforman la teoría del cambio según Weiss (1998). Rossi la denomina evaluación comprensiva (*comprehensive evaluation*¹²), ya que pretende abarcar un espectro amplio de la realidad del programa. Con relación a la lógica que subyace identifica tres tipos relaciones causales:

- 1) *Hipótesis causal*: son las relaciones entre los problemas a solucionar con las causas que los han creado.
- 2) *Hipótesis de la intervención*: establece la relación entre la acción del programa con los cambios en las causas.
- 3) *Hipótesis de la acción*: donde se vincula el programa con los cambios en el problema original (Fitzpatrick, 2009).

Críticas a la perspectiva de procesos

Se destacan cuatro bloques de críticas relacionadas con la programación, gestión o evaluación a través de procesos:

- La preocupación por ejecutar los procedimientos correctamente puede hacer perder el foco sobre los objetivos finales de toda la intervención. Intentando prevenir este problema, algunas iniciativas administrativas insisten en la necesidad de que los procesos tienen que estar orientados a los resultados, prueba de que esto no tiene por qué ser lo normal.
- A la gestión por procesos se le asocia otro problema, la pérdida de responsabilización de los profesionales. En un proceso se corre el riesgo de que cada actor se ocupe de su tarea asignada sin que se sientan responsables del producto final. A cada cual se le juzgaría por haber hecho bien su parte, aunque todo el programa fuera ineficaz nadie se sentiría directamente interpelado. Realmente esta crítica guarda similitud con el concepto de alienación hegeliano o marxista, donde lo producido es enajenado de su agente creador.
- Otro riesgo de la programación por procesos es que se caiga en un excesivo, pormenorizado e inoperante detalle de los procedimientos.

Se pueden crear procesos cada vez más complejos que pierden la flexibilidad necesaria para responder a las diversas contingencias que van apareciendo. En esta crítica se ha apoyado uno de los principales argumentos a favor de la gestión por objetivos. La óptica objetivos/resultados da toda la libertad a quien implementa (precisamente porque no define nada sobre cómo actuar). No obstante, las propuestas de modernización de la administración pública son conscientes de la tendencia a la complicación de los procedimientos de gestión e intentan buscar la simplificación de la acción.

- Otro de los elementos a tener en cuenta en el análisis de procesos es que no son abstracciones que se ejecutan en el aire, se requiere de unos recursos para poder implementarlos: trabajadores, alianzas, medios, materiales, etc., que no necesariamente se recogen en los procesos. Todos los elementos necesarios para la puesta en funcionamiento de un servicio o programa es lo que se puede denominar elementos estructurales.

3.1.3. Dimensión estructural. Los recursos, la organización y otros elementos estructurales

Por *estructura* se puede entender «la organización relativamente estable de distintos tipos de recursos para alcanzar los fines del proyecto» (Cohen y Franco, 1993: 94). Bajo esta definición se pueden englobar múltiples aspectos de un programa, pero todos ellos son elementos que se necesitan para que los procesos puedan funcionar. Se han escogido tres operacionalizaciones de la estructura que pueden resultar esclarecedoras para entender el concepto:

La primera de ellas son las claves enunciadas por Weiss para poder identificar las cuestiones relevantes de los «*inputs*, recursos y entorno del programa» (Weiss, 1998: 193), a saber:

- El auspicio organizacional.
- El presupuesto.
- Naturaleza del *staff*.

- La gestión.
- La localización.
- El plan de actividades.
- Los métodos de prestar el servicio.
- La filosofía o principios de operación.
- Los propósitos.

El segundo enfoque que se presenta es el modelo de excelencia TQM de EFQM¹³ (MAP). En esta propuesta se definen una serie de elementos previos a la acción que se pueden considerar estructurales, reciben el nombre de agentes facilitadores (en lo que también estarían incluidos los procesos):

- Liderazgo: desarrollo de valores y sistemas para el logro del éxito. Los líderes son capaces de reorientar a la organización en situaciones de cambios y arrastrar tras ellos al resto de personas.
- Política y estrategia: implantación de la misión, visión y el desarrollo de una estrategia centrada en los grupos de interés, teniendo en cuenta el mercado y el sector donde operan.
- Personas: los integrantes de la organización, ya sean individuos o grupos.
- Alianzas y recursos: planificación y gestión de las alianzas externas con otras entidades, con los proveedores y los recursos internos; todo ello al servicio de la política y estrategia.

El tercer enfoque es una propuesta de evaluación institucional (Lusthaus, Anderson y Murphy, 1995) que señala como aspectos clave de la capacidad institucional los siguientes elementos:

- Estrategia de liderazgo.
- Recursos humanos.
- Otros recursos clave (infraestructura, tecnología y finanzas).

- Gestión del programa.
- Gestión de procesos.
- Alianza interinstitucionales.

Todos estos elementos tienen en común el ser condiciones previas y necesarias para la ejecución de los procesos. El peligro del análisis estructural es que se convierta en un repaso de componentes sin identificar aquellos elementos que realmente son relevantes para la marcha del proyecto. Al analizar la estructura debe buscarse la vinculación entre los elementos estructurales y la ejecución de los procesos lo que producirá, si la teoría es correcta, los resultados esperados por la organización.

Con la estructura ocurre como con los procesos y los resultados, se puede programar, analizar y evaluar de forma aislada e independiente a las demás dimensiones de la intervención. Pero el análisis se enriquece si se incorpora como una tercera dimensión a la evaluación de resultados y de procesos. La vinculación de las tres dimensiones responde a una visión mecanicista de la organización donde todos los elementos funcionan como piezas que conforman un gran sistema para la producción de resultados.

3.1.4. La perspectiva sistémica

La vinculación entre procesos y resultados o la integración de diversas dimensiones de un programa, lo que Rossi ha llamado *evaluación comprensiva*, es probablemente «el avance más evolucionado y sofisticado en evaluación hasta la fecha» (Greene, 1999: 164).

El enfoque sistémico entiende el programa o servicio como un conjunto de elementos articulados e interdependientes entre sí que conforman un sistema. Las relaciones medios-fines recorren el sistema desde su inicio hasta los resultados finales implicando a todos los elementos de la intervención. La variación en uno de los elementos afecta al resto de los componentes. Por ejemplo, la reducción de los recursos asignados a un programa tendrá repercusiones en la ejecución de los procesos y, por lo tanto, en los resultados finales obtenidos. Para el buen funcionamiento del programa se requiere del funcionamiento correcto de todos sus elementos.

La especificación de las dimensiones de una intervención y las relaciones entre sí se denomina *teoría de la acción*, del programa o también *modelo lógico*. Se han desarrollado variados sistemas para identificar los modelos lógicos de las intervenciones. Por citar algunas propuestas, en 1966 Donabedian desarrolló un concepto integral de la calidad hospitalaria. Considera que «está influida por diversos elementos relacionados armónicamente y los agrupa en tres niveles: estructura física de la organización, proceso desarrollado para prestar a la asistencia y el resultado obtenido» (Martínez del Olmo, 2000: 274). Stufflebeam y colaboradores diferenciaron entre contexto, *inputs*, procesos y productos (CIPP¹⁴). Con estas dimensiones pretenden abarcar toda la secuencia clave para el desarrollo de una intervención (Stufflebeam y Shinkfifeld, 1995). En 1967 Robert Stake definió un modelo que proponía recoger información sobre los antecedentes, las transacciones y los resultados del programa para poder comprender y establecer juicios.

Una de las propuestas más referenciadas actualmente es la de la W.K. Kellogg Foundation (2004), las dimensiones básicas de análisis propuestas se resumen en el Gráfico 8.

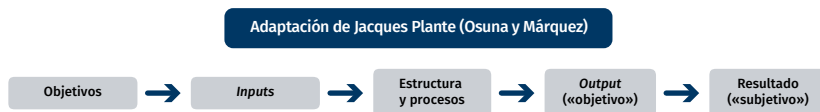
Gráfico 8: El modelo lógico



Fuente: Elaboración propia a partir del gráfico incluido en W.K. Kellogg, figura 1: *The Basic Logic Model*.

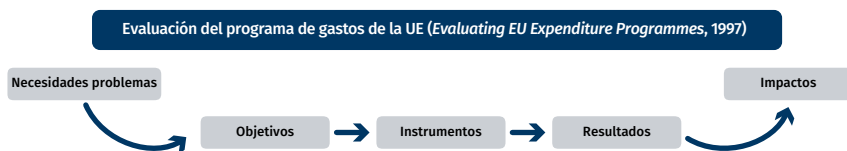
Otras propuestas teóricas como la de Klages y Haubner (Löffler, 1996: 31) discriminan entre resultados «subjetivos» y «objetivos» manteniendo la cadena causal (ver Gráfico 9).

Gráfico 9: Propuesta de Klages y Haubner



Fuente: Elaboración propia de Löffler.

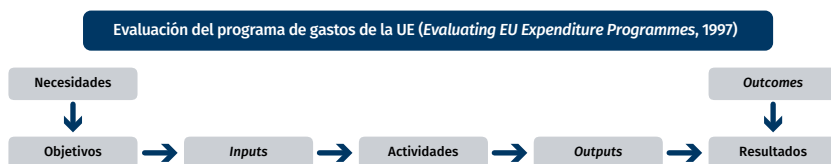
Gráfico 10: Adaptación de Jacques Plante



Fuente: Elaboración propia a partir de Osuna y Márquez.

La Comisión Europea trabaja con varias propuestas de modelos lógicos, como con la anterior de Plante (ver Gráfico 10), aunque en la guía *Evaluating EU Expenditure Programmes*, Nagarajan y Vanheukelen (1997) formulan el siguiente esquema (ver Gráfico 11).

Gráfico 11. Evaluación del Programa de Gastos de la UE

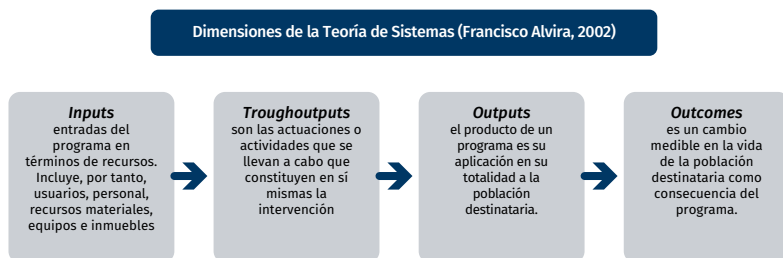


Fuente: Elaboración propia a partir de *Evaluating EU Expenditure Programmes*

El Observatorio Europeo de Drogas y Toxicomanías propone utilizar el modelo lógico propuesto por la Universidad de Wisconsin, que diferencia entre situación, *inputs*, *outputs* y *outcomes* (www.uwex.edu).

En la literatura española se encuentran varias propuestas como la de Paloma López de Ceballos (1989), en la que se diferencia entre finalidades, estructuras, métodos y entornos. Francisco Alvira (2002) expone cuatro dimensiones utilizando la terminología de la teoría de sistemas (ver Gráfico 12).

Gráfico 12: Dimensiones de la teoría de sistemas



Fuente: Elaboración propia a partir de Alvira (2002).

Todas estas propuestas son solo una muestra de la profusión de estrategias para identificar los modelos lógicos. Lo curioso es que varían poco en la conceptualización de lo que debe tenerse en cuenta en un programa u organización. Todas las visiones tienen mucho en común, lo que evidencia la existencia de un paradigma dominante sistémico en el análisis de las organizaciones (Blas, 2005: 74). En la práctica, la gran mayoría de las instituciones se basan en un modelo causal genérico de *inputs-process-outputs-outcomes* (Rogers *et al.*, 2000: 6).

Se haya hecho explícito o no, cualquier programa está proponiendo un determinado modelo teórico de cómo cambiar la realidad. En la medida en que la lógica en la que se apoya una intervención es una teoría, pueden existir otras alternativas, y todas ellas pueden ser validadas. Tal como dice Rein, un programa es una teoría y la evaluación su test (Weiss, 1998: 55).

La evaluación orientada por el modelo lógico dirige las reflexiones de mejora hacia los pasos concretos en la cadena causal. Esto permite conocer y consolidar cadenas causales exitosas, construir conocimiento sobre la intervención o cambiar aquellos supuestos tanto de la implementación como de los resultados que no están funcionando.

3.2. Múltiples perspectivas sobre las teorías que subyacen en una intervención

Todo este esquema funcionaría muy bien si existiera una única teoría del programa, tal como defiende el enfoque objetivista o también llamado

aproximación científica (Chen, 1990). La idea básica de los «objetivistas» es que la teoría de la intervención es una y ajena a la visión y percepción de los diferentes actores implicados. Pero hay varios aspectos incómodos en esta aproximación, el primero es que la teoría del programa declarada (a través de la programación o la documentación oficial) en muchas ocasiones no suele estar establecida en términos racionales ni tampoco está uniformemente aceptada por los diferentes actores. Pueden existir otras teorías, algunas de ellas explícitas pero muchas implícitas (Weiss, 1998: 55). El segundo aspecto es que puede haber una distancia entre lo programado y lo real, o dicho de otra forma, entre lo que se dice y lo que se hace.

El hecho de que haya una teoría explícita no implica que sea la versión objetiva. Como ya se ha comentado, la documentación suele representar la visión normativa. Monnier postula que, para formular las hipótesis relativas a los efectos producidos por un programa, lo que denominamos una teoría de acción, es preciso adoptar un determinado punto de vista (Monnier, 1992: 142).

Por tanto, el concurso de los diferentes actores implicados es un elemento crucial en la construcción de teoría o teorías para que sean fieles a la situación real.

3.2.1. Fuentes para obtener pistas y sugerencias para construir la teoría

- Realización de entrevistas a actores clave: gerentes y grupos de interés, extrayendo las presunciones, esquemas y expectativas con relación al programa. En esto coinciden gran parte de los autores.
- Revisión de la documentación del programa: Wholey recomienda tener en cuenta los documentos del programa, programación, legislación, ordenanzas internas, guías, justificaciones presupuestarias e informes de seguimiento o informes de logros (Chen, 1990: 66). En verdad, la documentación escrita se considera la plasmación de un determinado punto de vista, el institucional.
- Observar el programa: la observación puede ir dirigida a ver dónde realmente se gastan las energías (Weiss, 1998: 55; Lipsey y Pollar cit. Rogers *et al.*, 2000: 7). Se aconseja prestar más atención a lo ejecutado que a lo planeado.

- Revisión de la literatura en investigación y evaluación: la comprensión del fenómeno o de otras evaluaciones puede ayudar a entender la teoría.
- Ideas provenientes del propio equipo evaluador: diversos autores conceden legitimidad al equipo para proponer aspectos que haya identificado acerca de la teoría (Rogers *et al.*, 2000: 7).

3.3. Modelo lógico con perspectiva sistémica

La información recogida de las diferentes fuentes puede organizarse bajo algún modelo lógico. En el apartado «Perspectiva sistémica» se han visto algunos ejemplos. Wholey propone utilizar un esquema con los siguientes elementos: recursos, actividades, resultados y establecer las relaciones entre ellos, a lo que llama «diseño del programa».

Rossi y Chen proponen partir de los documentos del programa como la vía principal para detectar la teoría del programa confiando en los conocimientos y la experiencia de la evaluadora. Una vez elaborada esta primera aproximación, la contrastan con los puntos de vista de los actores clave. Si se dan discrepancias, propician una negociación entre los diferentes actores. La finalidad es lograr un consenso, «el entendimiento natural es mejor que lo obtenido por solo una parte» (Chen, 2000: 71).

Una forma de proceder es que las evaluadoras pueden construir uno o varios modelos lógicos tentativos que se discuten con los actores clave esperando encontrar confirmaciones o diferencias. A esta combinación de la propuesta «científica de la evaluadora» y las aportaciones «subjetivas de los actores» le han llamado «enfoque integrado».

Cuando los diferentes actores están de acuerdo sobre un aspecto, no hay mayor problema, esa dimensión del modelo queda confirmada. Pero en el caso de que haya discrepancias, hay que buscar consensos o al menos saber con claridad en qué no se está de acuerdo.

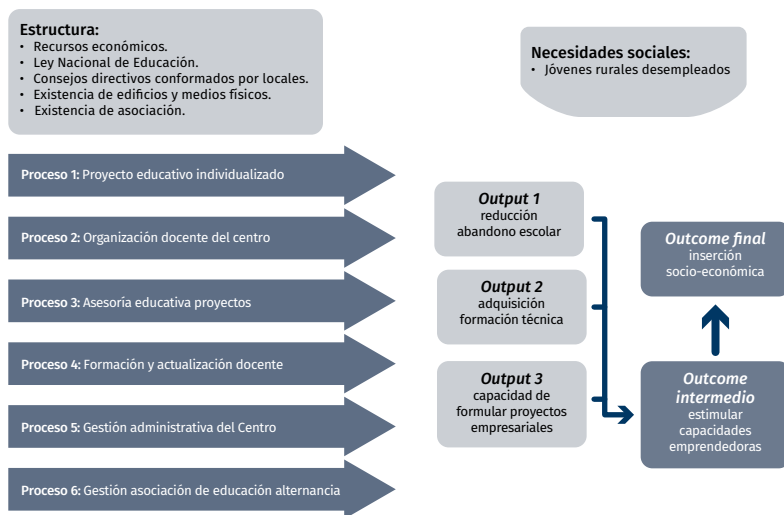
Wholey argumenta que el principal propósito de la construcción de la teoría es clarificar y desarrollar las teorías subyacentes de los diferentes actores implicados aunque estas sean diametralmente opuestas. Las evaluadoras no deberían confiar en su propio conocimiento o experiencia

para construir modelos. La teoría tiene que emerger de los datos más que de estructuras o hipótesis previas tanto de las evaluadoras como de otros actores (Chen, 2000: 67; Weiss, 1998: 76).

3.4. Graficar la teoría

Después de que las diferencias encontradas sean resueltas, pactadas o explicitadas, se completa formalmente una teoría o teorías del programa. Los diferentes componentes se presentan en un gráfico que muestre las cadenas causales (Rogers *et al.*, 2000: 7). Tal como citan estos mismos autores se pueden dibujar a través de una serie de «cajas» que representen las dimensiones de *inputs*, procesos, *outputs* y *outcomes* con flechas que las conecten entre sí. No es necesario especificar qué proceso lleva a qué output. Las relaciones entre los diferentes componentes serán exploradas en análisis posteriores. En el gráfico 13, se muestra un ejemplo simplificado de modelo lógico para un proyecto de educación en alternancia de la fundación CODESPA.

Gráfico 13. Ejemplo simplificado del proyecto Educación en Alternancia, CODESPA



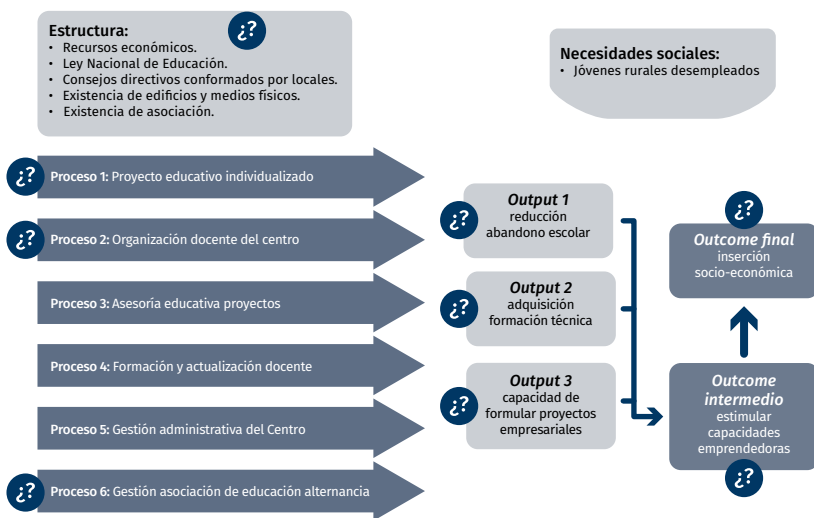
Fuente: Elaboración propia.

3.5. Determinación de los aspectos a evaluar

El modelo obtenido, la teoría, es usada como un marco para «determinar las variables que deben ser medidas en la evaluación» (Davidson, 2000: 18). Por tanto, se pueden identificar, de cada dimensión (resultados, procesos y estructura), los aspectos clave de la lógica que podrían ser tenidos en cuenta y medidos.

Las cuestiones deben cubrir toda la lógica de la intervención para después poder comprobar las relaciones causales que asume el programa. Si no se obtiene información de los aspectos clave, se corre el riesgo de no poder analizar la congruencia de la teoría.

Gráfico 14: Cuestiones clave sobre diferentes aspectos del proyecto



Fuente: Elaboración propia.

Carol Weiss (Weiss, 1998: 75) concreta una propuesta con cinco categorías de indagación:

- a) *Cuestiones sobre los resultados finales:* para identificar los resultados, una fuente de entrada es la «declaración de objetivos», pero existen otras fuentes (*staff*, clientes, etc.), además de procurar captar los objetivos no previstos. Davidson (2000: 22) aconseja la búsqueda libre de objetivos

para procurar recoger todos los cambios producidos por el programa. No hay que olvidar que por muy bienintencionada que sea una acción puede no tener efectos o incluso estos pueden ser contraproducentes.

- b) *Cuestiones sobre los procesos*: identificar los aspectos más importantes de los procesos y fases clave.

Además de saber si se están ejecutando de acuerdo a lo establecido.

- c) *Indagaciones acerca de si los resultados son atribuibles al programa*.

- d) *Indagación sobre si existen conexiones entre los procesos y los resultados*.

- e) *Cuestiones que facilitan las explicaciones sobre diferentes aspectos*: no solo se trata de saber qué es lo que ha pasado, sino cómo y por qué. También si los acontecimientos y resultados han sido producidos por las relaciones causales de la teoría del cambio.

3.6. Métodos y técnicas

Igual que ocurría en el modelo por criterios, las cuestiones tienen que ser contestadas a través de indicadores con sus respectivas fuentes de recogida de información (salvo en el caso de las técnicas cualitativas, que no requieren de indicadores). La evaluación basada en la teoría no tiene por qué estar adscrita a ningún tipo de técnica, puede ser abordada a través de técnicas cualitativas y cuantitativas. La combinación de técnicas y métodos es algo ya totalmente admitido por la práctica y la teoría de investigación (Serrano *et al.*, 2009). Davidson (Davidson, 2000: 22) propone que las fuentes o personas a entrevistar no sean las mismas que las que han participado en los pasos previos de construcción del modelo.

Una vez que se tiene la información de las diferentes dimensiones, la finalidad es estudiar las relaciones causa-efecto entre los *inputs*, los procesos y los resultados (Shadish y Leviton, 1991: 389). Es un caso específico de evaluación comprensiva tal como define Rossi, ya que busca articular e integrar las diferentes dimensiones del programa.

Uno de los métodos más extendidos para ver las relaciones entre las dimensiones es la modelización estadística a través de las ecuaciones lineales. Las correlaciones entre los diferentes elementos, controlando las posibles variables intervinientes, se utilizan como estimador de las relaciones causales. Es decir, la variación en los indicadores de la implementación se relaciona con la variación en los *outcomes* o las variaciones en los elementos estructurales con las variaciones con los procesos, asumiendo que si esto es así hay una relación causal entre dichas dimensiones.

Uno de los problemas de los sistemas de ecuaciones lineales es que en la práctica es una técnica muy exigente. Se necesita información cuantitativa de todas las variables que se quiere que estén en el sistema, además, todas deben ser independientes entre sí y cumplir una serie de criterios estadísticos. En el caso que se consiga tener la información y construir modelos con validez interna, en muchas ocasiones las varianzas explicadas suelen ser muy bajas, por lo que pierde capacidad explicativa.

Se pueden construir modelos menos ambiciosos a través de medidas de asociación, como el estadístico chi cuadrado y tablas de contingencia. La dificultad de esta opción es controlar múltiples variables intervinientes lo que incrementa el riesgo de sesgo.

Los experimentalistas entienden que la correlación o la asociación no son suficientes para establecer una atribución causal. Entre otras razones porque puede haber variables influyentes que se escapan del modelo y porque no hay validación contrafactual que pueda descartar otros modelos causales alternativos (Davidson, 2000: 23).

Thomas Cook critica la evaluación orientada por la teoría por su predisposición a utilizar modelizaciones causales y porque puede disuadir en la utilización de modelos experimentales. Aunque autores como Rubin o el mismo Cook creen que es posible evaluar basándose en la teoría, pero a través de métodos experimentales (Cook, 2000: 34).

Existen otras aproximaciones metodológicas de corte naturalista, principalmente apoyadas en técnicas cualitativas. Se pueden reconstruir las teorías a través del discurso, las historias de vida o presentación de

casos individuales, tal como lo hace Robert Stake (2006). No se busca una generalización de tipo cuantitativo, sino ayudar a los diferentes actores a comprender mejor que es lo que ocurre en su programa.

En cualquiera de los casos, el propósito de los métodos es reconstruir «un modelo causal apoyado en varias fuentes para entender cómo una intervención específica (o programa) está intentando lograr sus productos y resultados» (Stame, 2004).

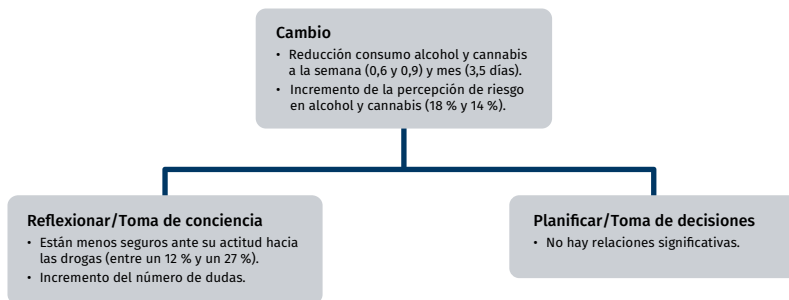
3.7. Análisis

El objetivo de la evaluación de la teoría es comprobar el modelo lógico establecido. Una vez que se tiene la información, se establece una secuencia de análisis que indague sobre si el programa funciona o no y a qué es debido (Rogers *et al.*, 2000: 10). La estructura del análisis guarda relación con la estructura de la indagación y de las preguntas. La secuencia puede ser la siguiente:

a) Análisis de la jerarquía de objetivos

Se estudia el grado de alcance de los objetivos y el grado de correlación existente entre los diferentes niveles o pisos de objetivos. En un programa de prevención de consumo de drogas (Ligeró *et al.*, 2010), teníamos dos pisos de objetivos. Obtuvimos resultados para cada uno de los objetivos de forma aislada, tal como se muestra en el Gráfico 15:

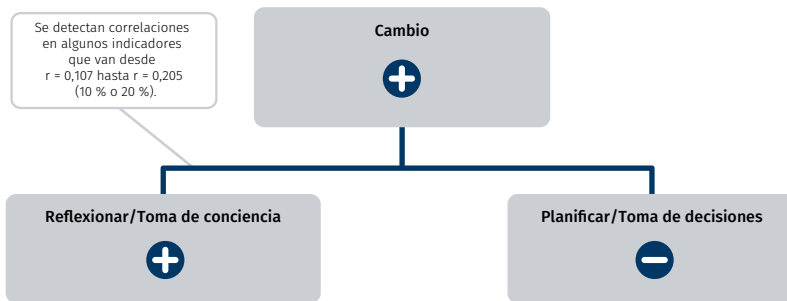
Gráfico 15. Análisis de la Teoría del Programa de un programa de prevención de consumo de drogas (Ligeró *et al.*)



En uno de los objetivos intermedios «Planificar/Toma de decisiones», no se detectaron cambios significativos. En el resto de los objetivos sí.

El siguiente paso fue analizar las relaciones entre los dos niveles de objetivos. En el ejemplo que se está utilizando, se analizó la asociación entre reflexionar/toma de conciencia y cambio con relación al consumo. Resumimos los resultados en el Gráfico 16.

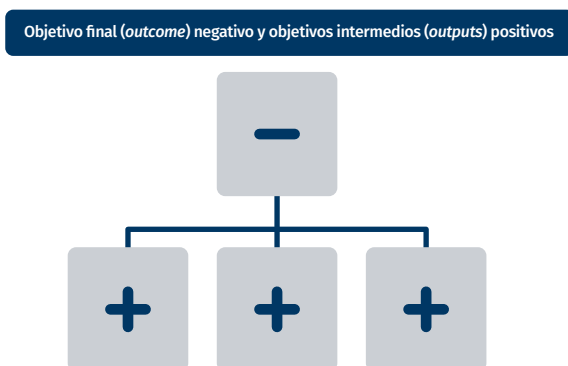
Gráfico 16. Análisis bivariados entre indicadores del análisis de un programa de prevención de consumo de drogas (Ligero *et al.*)



Es decir, se detecta una relación entre el objetivo intermedio «reflexión» y el objetivo final «cambio»: una variación en el objetivo intermedio produce una reducción del consumo de alcohol y cannabis al mes entre un 10 % y un 20 %. En este caso hay indicios que permiten establecer una relación lógica entre el objetivo intermedio y el final.

Pero puede haber otras posibles situaciones. En el Gráfico 17, se resumen los resultados de los indicadores, el símbolo negativo es que no se ha alcanzado y el positivo que sí.

Gráfico 17. Esquema sintético de resultados



Fuente: Elaboración propia.

Los objetivos intermedios han sido alcanzados, pero no hay cambios en el objetivo final. En esta teoría del programa, tal como está planteada, no hay una relación causal (sí hay un control efectivo de otras posibles variables intervinientes). Puede ser que con el concurso de otros objetivos intermedios se logre el cambio, pero tal como está planteado no tiene sentido continuar ejecutando el programa, ya que no hay conexión causal entre los dos pisos.

Estos dos ejemplos ilustran sobre las posibilidades del análisis de la jerarquía de objetivos. Se puede testar y validar la teoría en la que se apoya el programa e incluso conocer la fuerza de la relación. Es decir, se podría llegar a saber no solo cuánto influye el programa en la realidad, sino cuánto influye cada uno de los resultados esperados en el logro del objetivo específico.

b) Análisis del cumplimiento de procesos y la relación con los resultados

Lo que se pretende básicamente en este apartado es conocer, tal como argumenta Cronbach, «qué procesos contribuyen a los resultados esperados» (Ballart, 1992: 340). Las cosas que hace el programa, los procesos, teóricamente deben conducir al logro de los resultados. Si no se consiguen los resultados (y se han propuesto métodos para desechar que no sea debido a causas externas), hay que pensar que los procesos no se han ejecutado correctamente o que han sido mal diseñados.

La explicitación de las dos dimensiones (resultados y procesos) permite avanzar en el análisis de la congruencia teórica de la intervención. Por ejemplo, en el supuesto de que los procesos se hayan ejecutado correctamente (e insisto, en que se consigan controlar las variables externas), se deberían lograr los resultados previstos si la teoría de la implementación es correcta (ver Gráfico 18).

Gráfico 18. Procesos y resultados positivos



Fuente: Elaboración propia.

Y a la inversa, si los procesos están mal diseñados o se han ejecutado indebidamente, es lógico pensar que no se lograrán los resultados (ver Gráfico 19).

Gráfico 19. Procesos y resultados negativos



Fuente: Elaboración propia.

Tanto Bickman, Lipsey como Weiss destacan que este tipo de análisis es útil para distinguir entre los fallos de la teoría del programa (jerarquía de objetivos) o los fallos en la implementación (procesos) (Rogers, 2000: 8). Vinculando la información de esta manera se podrá saber si el fallo está en la lógica de resultados, en la lógica de procesos o en la implementación. Pero no solo de una forma genérica, si la información ha sido exhaustiva, se podrá definir en qué fase o fases de los procesos se está fallando.

c) Análisis de los elementos estructurales

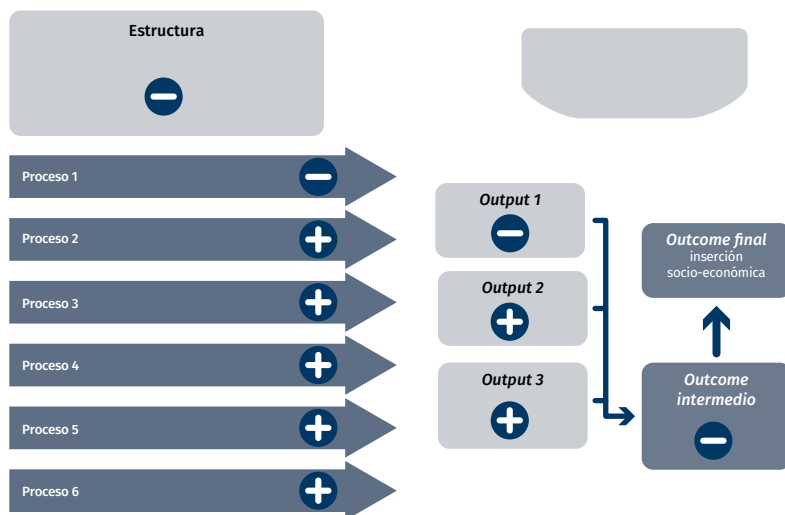
Por último, si un proceso no se ejecuta de acuerdo a lo definido o tiene problemas de implementación, las causas hay que buscarlas en los

elementos estructurales, que son los medios necesarios para poner en funcionamiento la intervención. Pudiera ser que los resultados no se consigan porque los procesos tampoco se han ejecutado correctamente, ya que no contaba con los medios adecuados. Si se ha previsto en el diseño de la evaluación, se podrá saber qué factores estructurales provocan la mala implementación.

d) Análisis de la cadena causal

Lo que se intenta poner de manifiesto con los análisis anteriores es conocer la cadena causal de medios y fines que recorre el programa para saber dónde ha fallado este o cuáles son las claves del éxito (Weiss, 1998: 66). Una vez que se conocen las relaciones causales, se sabe dónde aplicar correcciones o reforzar actuaciones. Es un modelo centrado en la comprensión de lo que se hace y sienta unas bases concretas para orientar la mejora.

Gráfico 20. Posibilidades de análisis: positivos y negativos



Fuente: Elaboración propia.

3.8. Ventajas de la evaluación de la teoría

- El ejercicio de *explicitar la teoría o las teorías* en sí mismo es positivo, ya que obliga a pensar o repensar la lógica de la intervención. Es un ejercicio que incrementa la racionalización de la intervención. Incluso si la «racionalización» del programa no es la correcta, la explicitación ayuda a la intervención. Weick recoge la historia de una unidad de reconocimiento perdida tres días en los Alpes en una tormenta de nieve. Finalmente, encontraron el camino de vuelta al campamento y se salvaron con la ayuda de un mapa. Un mapa que más tarde descubrieron que era de los Pirineos, no de los Alpes. Tal como dice el autor, «tener un preciso mapa de la zona puede ser menos importante que tener algún mapa que traiga orden al mundo y empuje a la actuación» (Rogers, 2000: 54).
- Contribuye a *reducir el disenso y las contradicciones* entre los diferentes actores implicados con relación al programa. Se favorece la búsqueda de denominadores comunes sobre la acción. En el caso de que no resuelvan las discrepancias, se contribuye a explicitar el disenso.
- El desarrollo de la evaluación basada en la teoría provoca un *incremento de la comprensión* de los mecanismos por los cuales el programa produce sus efectos (Davidson, 2000).
- *Validar el marco teórico*: En el caso de que los datos sean positivos y la teoría congruente, se está ante una lógica de intervención válida al menos en el contexto donde se ha evaluado (aunque hay que tener en cuenta las críticas de los experimentalistas ante este tipo de atribución). Se puede sistematizar la experiencia y difundir tanto en ámbitos académicos como de acción, aportando una nueva información en el proceso de construcción de conocimiento. Otros actores podrán utilizar ese modelo cuando intenten solucionar esos problemas en otros lugares.
- *Modificar/ajustar el marco teórico*: De las conclusiones de la evaluación se extraen indicios razonables que justifican la modificación tanto de la lógica de los objetivos como de los procesos o de los elementos estructurales. Se puede saber dónde la cadena o las cadenas causales no están funcionando o se ha roto. Por lo tanto, la evaluación de la

teoría se dirige con precisión hacia los aspectos mejorables e identifica de forma concreta lo que hay que cambiar de los resultados, de la implementación o de los medios.

- *Cambiar la lógica de la intervención:* Una de las posibles conclusiones es que el marco teórico del programa no funcione. Muchos programas tienen impacto cero debido a unas presunciones ineficaces. La evaluación puede ayudar a legitimar y fundamentar un cambio de orientación parcial o total.
- *Toma de conciencia:* El que los diferentes actores implicados incrementen su conocimiento sobre el programa, conozcan más, cambien la percepción y, en definitiva, tomen conciencia sobre lo que ocurre es un fin en sí mismo. No solo de algunos enfoques de evaluación, sino también de otras metodologías, como la Investigación-Acción Participativa. Se parte del principio de que los cambios tienen que ser endógenos y de que no hay cambio sin un convencimiento previo de que esto es necesario. Por eso los procesos participativos buscan que los propios actores lideren el proceso de evaluación o investigación, para que a partir de una toma de conciencia real sobre la situación se busquen acciones de consolidación o de reforma. Patton y Saville afirman que se contribuye mejor al desarrollo de la organización cuando ayudan a que el estudio se convierta en una autoevaluación (Stake, 1967: 85). La teoría del programa, por su ordenación lógica de la práctica, ayuda a los diferentes actores a comprender y dotar de sentido a los datos procurando incitar la toma de conciencia.

3.9. Debilidades y críticas

- Existe una confianza excesiva sobre las posibilidades reales de validar la teoría de la acción (Davidson, 2000). Existen dudas sobre que los métodos estadísticos permitan sacar conclusiones sobre la atribución causal. Tomás Cook (2000) critica que la modelización permita desechar otras teorías alternativas.
- La gran influencia del paradigma constructivista en la elaboración de teorías y del peso otorgado a los diferentes puntos de vista de los actores implicados puede ser acusado de modelo «*subjetivista*» con

la consiguiente pérdida de credibilidad. De hecho, Rossi declara que el *staff* u otros actores no son perspicaces para identificar la cadena causal, siendo mejor que lo hagan los científicos, entendiendo por ello a las evaluadoras.

- La crítica opuesta es mantenida por Stufflebeam (2001) quien, por el contrario, cree que el hecho de que la evaluadora extraiga o negocie con la teoría la *sesga y usurpa la responsabilidad del diseño del staff*.
- Las discrepancias entre los actores a la hora de definir la teoría puede suponer un bloqueo del proceso de evaluación o un desgate y pérdida de energías en procesos de negociación que pueden enmascarar conflictos latentes de la organización previos a la evaluación.
- La *generalización de los hallazgos* es cuestionada porque está muy condicionada por el contexto. La validez de una teoría en un lugar puede no serlo en otro. No obstante, determinadas aproximaciones defendidas por Cronbach no conceden mucha importancia a este tipo de validez, es más importante la contribución que logra el proceso de evaluación en la toma de conciencia de los actores implicados.

4. Conclusiones finales

Utilizar un método u otro de evaluación no es indiferente, tiene consecuencias que incluso pueden llegar a resultados antagónicos. Las dos propuestas presentadas en este texto, la evaluación por criterios y la evaluación de la teoría del programa, están respaldadas institucionalmente por la teoría y validadas en la práctica, pero tienden a generar diferentes productos y están dirigidas a diferentes propósitos evaluativos.

María Bustelo (2010) recopila tres propósitos para la realización de evaluaciones. Una adaptación de ellos se muestra en el siguiente esquema (ver Tabla 7):

Tabla 7. Propósitos de la evaluación

| Propósitos | Concepto e ideas |
|---|---|
| A) Ilustración, comprensión (<i>enlightenment</i>) | Ilustrar, aclarar, dar luz, comprender qué está pasando y cómo funciona el programa. La evaluación suele producir un cambio, dota de nuevos significados sobre la realidad del programa y de su intervención. |
| B) La retroalimentación o mejora (<i>improvement</i>) | Es una manera de recibir « <i>feedback</i> » sobre las acciones, intervenciones o programas que se llevan a cabo, un modo de mejorar y progresar. La evaluación puede ser un proceso o de aprendizaje de la propia evaluación. |
| C) El control de responsabilidades y rendimiento de cuentas (<i>accountability</i>) | «Rendir cuentas» en la acepción más flexible del término, no se restringe exclusivamente a los aspectos económicos, sino también a la responsabilidad social y política, especialmente si las actividades llevadas a cabo se financian con fondos públicos ya que el contribuyente tiene derecho a saber en qué y cómo se está empleando el dinero público. |

La evaluación por criterios produce juicios de valor sobre el programa en cada uno de los criterios definidos. No requiere de una comprensión previa para emitir los juicios, sobre todo si el diseño está realizado de antemano como suele ser usual en la cooperación. La comprensión se puede dar, pero el modelo no está orientado hacia esa finalidad. Su funcionalidad se encuentra en el *accountability*, en el rendimiento de cuentas público.

La evaluación de la teoría del programa tiene como finalidad comprender las cadenas o secuencias causales que provocan los resultados. Su intención es saber por qué han ocurrido las cosas. Cuando se sabe el porqué es cuando se puede mejorar, proponiendo recomendaciones enfocadas y concretas que atañen a la totalidad de la intervención. Si se quiere enjuiciar después de la comprensión se puede, pero el modelo no está dirigido hacia el juicio, no lo necesita, ya que su finalidad principal es la ilustración (*enlightenment*) y la comprensión.

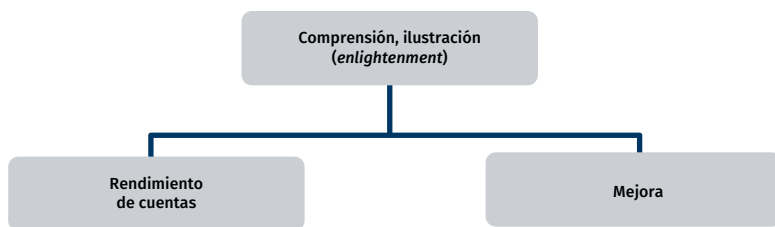
En mi experiencia profesional, he podido ver cómo el juicio perdía su importancia ante la fuerza y la significación de la comprensión. Si a los diferentes actores relacionados con el programa les anima una intención constructiva, lo que demandan es la comprensión del fenómeno para poder solucionarlo. Para explicar esto me gusta recurrir a un ejemplo de un estudiante: sus padres pueden recibir informes sobre si el niño es buen o mal estudiante, sobre sus capacidades, aptitudes, actitud y colaboración en el colegio u otros muchos criterios, pero estarán mucho más receptivos a un informe donde les expliquen concretamente las causas del mal rendimiento. Quizá la razón sea que el estudiante no sabe estudiar, nadie le ha enseñado, e invierte con tesón todas sus energías en acciones inútiles.

Después de saber esto, el enjuiciamiento en función de criterios pierde importancia. Ya se conocen las causas y se pueden buscar acciones de mejora, indistintamente de que se le haya calificado como buena o mala en cualquiera de los criterios.

Lo mismo ocurre con los proyectos de desarrollo, enjuiciar a través de criterios está bien, pero si nos anima una intención de mejorar lo que se está evaluando, pueden ser más útiles sistemas de evaluación centrados en la comprensión de lo que ocurre. Cuando se comprende y se toma conciencia de las dificultades que se tienen también se sabe cómo mejorar. El rendimiento de cuentas público puede concluir diciendo que es necesario mejorar, pero no saber exactamente cómo hacerlo.

La mejora de las intervenciones en busca de la calidad está más cómoda en las propuestas de evaluación comprensivas que en las centradas en el *accountability*. Eso me ha llevado a proponer una secuenciación entre los propósitos de evaluación, en la que aparece primero la comprensión y después el enjuiciamiento y la mejora.

Gráfico 21. Secuencia de los propósitos



Fuente: Elaboración propia.

Cada uno de estos propósitos tiene sentido en sí mismo, habrá que saber qué es lo que se busca con cada evaluación en concreto. Pero si lo que se quiere es la mejora, será más importante pensar en sistemas centrados en la comprensión como la evaluación de la teoría del programa para después poder deducir recomendaciones concretas. Si, por el contrario, lo que se busca es producir un rendimiento de cuentas público, un sistema adecuado es la evaluación por criterios.

Existe una literatura propia sobre el escaso uso y utilidad de las evaluaciones y también he escuchado a responsables políticos pensar en «obligar» a usar las evaluaciones. Creo que son síntomas de una distancia de la evaluación con las necesidades reales de clientes y otros *stakeholders*. Si lo que se quiere es mejorar con sistemas centrados en el *accountability*, no es extraño que no se usen las recomendaciones, ya que no se ha generado el producto que es previo al cambio, la comprensión. No hay cambio sostenible si los propios actores implicados no han comprendido lo que ocurre y tomado conciencia de la necesidad de cambiar.

Igual que ocurría en la definición de criterios *a priori*, los diferentes actores en el sector de la cooperación pueden tener intereses diferenciados y buscar los métodos de evaluación más adecuados a ellos. La OCDE y el CAD pueden proponer un modelo que busque el *accountability*, ya que sus prioridades y demanda de evaluación puede ser esa, lo cual, por otro lado, es legítimo. Pero los beneficiarios, técnicos y ONG pueden tener otros intereses y demandas hacia la evaluación que también son legítimas y que pueden no verse reflejadas del todo en el modelo de criterios.

La hipertrofia del modelo de criterios en la cooperación responde probablemente a su dependencia característica de los financiadores. Se ha desarrollado de tal manera que se está ante una propuesta metodológica «institucional». Es decir, la institución o instituciones definen qué método hay que utilizar, cuando normalmente esto suele ser atribución del campo técnico o académico. En otros sectores no es así, se sorprenden cuando se asoman a la cooperación. Es como si el Ministerio de Sanidad exigiera solamente evaluar con determinados criterios de eficiencia o eficacia, con técnicas experimentales o el Ministerio de Educación solo con el modelo CIPP de Stufflebeam. Recibirían merecidas críticas por limitar sus posibilidades, el ajuste a las necesidades de los diferentes actores y se sospecharía de las razones que les llevan a atreverse a tal imposición.

Teóricamente, no hay ninguna presión para utilizar el método de criterios. Según el CAD (2010), sus propuestas intentan proveer de una guía hacia las buenas prácticas. Además, se puede encontrar en sus textos una apertura en la que caben otros tipos de evaluación y concretamente, la evaluación orientada por la teoría. Así lo expone Nicoletta Stame (2004),

entendiendo que la evaluación de la teoría del programa forma parte de los métodos del CAD. En el punto 2.9 del documento citado del CAD, se dice que «la metodología se elabora de acuerdo con el enfoque elegido. (...) La metodología seleccionada debe responder a las preguntas de la evaluación utilizando evidencias creíbles» (CAD, 2010: 10).

Pero tal como lo definen Osset y Touza¹⁶, esto es contradictorio con lo que aparece antes en el punto 2.8. En él se especifica que en evaluación se aplican los criterios adaptados por el CAD referentes a la evaluación de la ayuda al desarrollo (CAD, 2010: 9). Aunque las consignas del CAD no sean vinculantes, no hay razón metodológica y técnica para no aplicar los consejos de una organización de la que España participa.

El peso de la institución (OCDE) y de la propuesta del CAD asfixian la necesaria flexibilidad metodológica y convierten cualquier otra propuesta que se salga de los criterios en arriesgada aunque puede estar mayoritariamente aceptada por la comunidad científica y sean «solo unos pocos quienes fijan estándares por adelantado» (Stake, 2006: 243).

En cualquier caso, los dos métodos son compatibles, se pueden aplicar ambos a la vez. Con esto se puede responder un contexto metodológico rígido y exigente y a la necesidad de sacar más partido a la evaluación. Aunque, si uno puede elegir, sería mejor primero comprender qué es lo que ocurre en el programa antes de que los juicios sumativos nos dificulten el proceso de mejora de los programas.

Notas

* Título original “Dos métodos de evaluación: criterios y teoría del programa: criterios y teoría del programa”, Madrid, CECOD, CEU Ediciones del Centro de Estudios de la Cooperación, Fundación Universitaria, San Pablo, 2011. Reproducción autorizada.

El autor agradece la revisión realizada por Óscar Franco Alonso para esta edición.

¹ Gran parte de lo que a continuación se va a narrar procede del texto *Operacionalización* (Ligero, 2010 [en curso]). Materiales del magister en Evaluación de Programas y Políticas Públicas, Universidad Carlos III de Madrid.

² Convención de género en la redacción. Dentro de las diferentes propuestas para no invisibilizar a ningún género, se ha optado en este texto por utilizar el femenino para

referirse a los y las evaluadoras, el resto de los actores serán citados en masculino.

3. En castellano, existe la voz mexicana *responsivo*: perteneciente o relativo a la respuesta (RAE, 2001).
4. Actor afectado o que afecta al programa o a la evaluación (Bustelo, 2010).
5. Dirección General de Planificación y Evaluación de Políticas de Desarrollo.
6. El término *primer trabajo de campo* no implica necesariamente viajar a terreno. Se pueden obtener las necesidades informativas a través de diversas técnicas y vías aunque implicará un trabajo de obtención, recopilación y tratamiento de la información.
7. Proyecto de la Fundación CODESPA.
8. Ver <http://www.fundacionlealtad.org/web/home>.
9. Por ejemplo, en el sistema de programación del Enfoque del Marco Lógico los objetivos intermedios se denominan directamente «resultados esperados».
10. Uno de los sistemas de programación por objetivos más extendidos es el Enfoque del Marco Lógico o el modelo ZOPP de la alemana GTZ. Se pueden tener en cuenta también los modelos de racionalización presupuestaria como el PPBS norteamericano, el RCB francés o su equivalente RFA de los países bajos (Monnier, 1992: 30).
11. Un proceso (del latín *processus*): Conjunto de fases sucesivas de un fenómeno natural o de una operación artificial (DRAE, 2001).
12. Se ha traducido *comprehensive* como comprensivo, ya que el término castellano comprender incluye la acepción de abrazar, ceñir, rodear por todas partes algo (DRAE).
13. European Foundation for Quality Management.
14. *Context, Inputs, Process y Products*.
15. Se realizaron correlaciones entre diversos indicadores controlando algunas variables externas.
16. Comentarios realizados por Iván Touza a los estándares del CAD. Grupo de trabajo: Métodos de Evaluación. Magister en Evaluación de Programas y Políticas Públicas, UCM.

Referencias

- Alvira, F. (2002). *Manual para la elaboración y evaluación de programas de prevención de abuso de drogas*. Madrid: Comunidad de Madrid, Agencia antidroga.
- Bachman, J.G., O'Malley, P., Schulenberg, J.E., Johnston, L.D., Bryant, A.L. y Merline, A.C. (2002). *The Decline of Substance Use in Young Adulthood: Changes in Social Activities, Roles and Beliefs*, Mahwa, NJ: Lawrence Erlbaum Associates.

Blas, F.A. (2005). Evolución histórica del estudio del comportamiento organizacional. En Gil, F. y Alcocer, C.M. *Introducción a la psicología de las organizaciones*. Madrid: Alianza Editorial.

Ballart, X. (1992). *¿Cómo evaluar programas y servicios públicos?* Madrid: MAP.

Becoña, E. (1999). *Bases Teóricas que Sustentan los Programas de Prevención de Drogas*, PNSD.

Bickman, L. (Ed.) (1987). Using program theory in evaluation. *New Directions for Program Evaluation*, 33. San Francisco, CA: Jossey-Bass.

Bustelo, M. (2001). La evaluación de las políticas públicas de igualdad de género de los gobiernos central y autonómicos en España: 1995-1999. Madrid: UCM. Tesis Doctoral. Publicación electrónica.

—(2002). ¿Qué tiene de específico la metodología de evaluación? En Bañon, R. (Eds.), *La evaluación de la acción de las políticas públicas*. Madrid: Díaz de Santos.

—(2006). The potential role of standards and guidelines in the development of an evaluation culture in Spain. *Evaluation*, 12: 437-453.

—(2010). *Materiales Magíster en evaluación* Madrid: UCM.

CAD (1995). *Principios del CAD para una ayuda eficaz. Manual de la ayuda al desarrollo*. Madrid: MundiPrensa/OCDE.

—*Estándares de Calidad para la Evaluación del Desarrollo* (2010). OCDE.

Cea D' Ancona, M.A. (1996). *Metodología cuantitativa. Estrategias y técnicas de investigación social*. Madrid: Síntesis.

Chen H-T. (1990). *Theory-Driven Evaluations*. California: Sage.

Cohen, E. y Franco, R. (1993). *Evaluación de Proyectos Sociales*. Madrid: Siglo.

Cook, T. (2000). The False Choice between Theory-Based Evaluation and Experimentation. En Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (Eds.) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluations, 87.

Davidson, E.J. (2000). Ascertaining Causality in theory-Based Evaluation. En Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (Eds.) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluations, 87.

Diccionario de la Real Academia Española (2011).

Fitzpatrick, J. (2009). *Ponencia Magíster Evaluación de Programas*. Magíster Evaluación de Programas, UCM.

Giménez, C. (2007). Ponencia presentada en el curso *Retos y Perspectivas de la inmigración: las segundas generaciones*. U. Carlos III de Madrid.

- Greene, J. C. (1999). *The inequality of performance measurements*. *Evaluation* 5: 160-172.
- Hintze, J. (2005). Instrumentos de evaluación de la gestión del valor público. En Echebarría et al. *Responsabilización y evaluación de la gestión pública*. Caracas: CLAD.
- House, E. (2006). Democracy and Evaluation. *Evaluation* 12 (1): 119-127.
- Leeuw, F.L. y Furubo, J.E. (2008). Evaluation Systems. What are they and why study them? *Evaluation* 14, 157-169.
- Ligeró, J.A. (2001). Propuesta de un Nuevo Criterio de Evaluación, la Cobertura. *Revista Española de Desarrollo y Cooperación*, N.º 8, primavera-verano.
- (2010). *Comprender la Evaluanda* Madrid: Magíster Evaluación, UCM.
- Ligeró, J.A., Franco, O., García, O. y Pariente, A. (2010). *Informe de evaluación del proyecto ¿Sales Hoy?* Madrid: Cruz Roja Española, Comunidad de Madrid.
- Löffler, E. (1996). *La Modernización del Sector Público desde una Perspectiva Comparativa. Conceptos y Métodos para Evaluar y Premiar*. La Calidad en el Sector Público en los Países de la OCDE. Madrid: Documentos INAP.
- López de Ceballos, P. (1989). *Un método para la Investigación-Acción Participativa*. Madrid: Editorial Popular.
- Lusthaus, C., Anderson, G. & Murphy, E. (1995). *Institutional Assessment: A Framework for Strengthening Organizational Capacity*. Ottawa: International Development Research Centre.
- MAE (Ministerio de Asuntos Exteriores). (2007). *Manual de Gestión de Evaluaciones de la Cooperación Española*. Madrid: MAE, Secretaría de Estado de Cooperación, DGPOLDE.
- (1998) *Metodología de Evaluación de la Cooperación Española*. Madrid: Ministerio de Asuntos Exteriores, Secretaría de Estado para la Cooperación Internacional y para Iberoamérica.
- MAP (Ministerio de Administración Pública). (2004). *Modelo EFQM de excelencia. Caso práctico para la Administración Pública*. Madrid: MAP.
- Martínez del Olmo, M.T. (2000). Evaluación de la calidad de los servicios sanitarios. Trinidad, A. *Evaluación y Calidad en las Organizaciones Públicas*. Madrid: INAP.
- Monnier, E. (1992). *Evaluación de la acción de los poderes públicos*. Madrid: Instituto de Estudios Fiscales.
- Nagarajan N. y Vanheukelen, M. (1997). *Evaluating EU Expenditure Programmes*. European Commission.
- Osset, A. (2010). *Diseño de una evaluación del sistema de gestión de evaluaciones de convenios firmados entre las ONGD y AECID*. Memoria de grado Magíster Evaluación de Programas y Políticas Públicas. Madrid: CSEG, UCM.

Rodríguez, L. (2008). *Evaluaciones Colaborativas*. Llumina Press.

Rogers, P.J. (2000). Causal Models in Program Theory Evaluation. En Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (Eds.) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluations, 87.

Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (2000). Program Theory Evaluation: Practice, Promise, and Problems. En Rogers, P.J., Hacsí, T.A., Petrosino, A. & Huebner, T.A (Eds.) *Program Theory in Evaluation: Challenges and Opportunities*. New Directions for Evaluations, 87.

Rutter, M., Giller, H. y Hagell, A. (2000). *La conducta antisocial de los jóvenes*. Madrid: Cambridge University Press.

Scriven, M. (1991). Beyond Formative and Summative Evaluation. En: Mclaughlin, W. & Phillips, D.C. (edit.). *Evaluation and Education: At quarter century*. National Society for the Study of Education.

Serrano, A., Blanco, F., Ligeró, J.A., Alvira, F., Escobar, M. y Sáenz, A. (2009). *La investigación multimétodo*. UCM.

Shadish, W., Cook, T. & Leviton, L. (1991). *Foundations of Program Evaluation*. Ca: Sage.

Stake, R.E. (1967). The countenance of educational evaluation. *Teachers College Record*, Volume 68, N.º 7, April.

—(2006). *Evaluación comprensiva y evaluación basada en estándares*. Barcelona: Grao.

Stame, N. (2004). Theory-based Evaluation and Varieties of Complexity. *Evaluation*, 10, 58-76.

Stufflebeam, D. (2001). *Evaluation Models*. San Francisco: New Directions in Program Evaluation.

Stufflebeam, D. y Shinkfield, A. (1995). *Evaluación sistemática. Guía teórica y práctica*. Barcelona: Paidós.

Weiss, C. (1998). *Evaluation*. New Jersey: Prentice-Hall.

W.K. Kellogg Foundation (2004). *Logic Model Development Guide*. Michigan: W.K. Kellogg Foundation.

Un discurso de políticas que refleja su marco*

Martin Rein y Donald Schon

El enmarque del discurso de las políticas

Mientras que en la primera parte del volumen se han examinado las experiencias nacionales con respecto a la relación que hay entre el conocimiento social y el Estado, la segunda parte aborda cuestiones acerca de la relación que existe entre ambas esferas. La primera cuestión que abordaremos es el carácter de los procesos de la política pública a los que se espera que contribuya la ciencia social. Este capítulo y el siguiente, de Majone, intentan conceptualizar el proceso de políticas en formas innovadoras, con objeto de aclarar la relación que guarda con la ciencia social orientada hacia las políticas.

En la práctica de las políticas existen obstinadas controversias que tienden a eternizarse, relativamente inmunes a toda resolución por referencia a los testimonios y que rara vez llegan a resolverse en forma definitiva. En el mejor de los casos quedan zanjadas durante un tiempo por obra de procesos electorales, tomas del poder o negociación. O, con los cambios ocurridos en un contexto mayor, simplemente pueden desaparecer por un tiempo, para reaparecer en alguna forma nueva.

Las trayectorias de estas controversias —podemos pensar en las disputas causadas por las armas nucleares, el bienestar o la condición de las mujeres— no deben interpretarse en términos de la ya familiar separación de cuestiones de valor y cuestiones de hecho, pues los participantes en

ellas interpretan los problemas de sus difíciles situaciones de políticas por medio de *marcos* en que se integran hechos, valores, teorías e intereses. Dadas las múltiples realidades sociales creadas por marcos en conflicto, los participantes no solo están en desacuerdo mutuo, sino también acerca de la naturaleza de sus desacuerdos.

Las controversias en torno a las políticas duraderas plantean la siguiente cuestión epistemológica: ¿cuál puede ser la base para resolver conflictos de marcos cuando los propios marcos determinan lo que cuenta como prueba, y cómo se ha de interpretar la prueba? Como respuesta a esta situación proponemos en este capítulo una epistemología empírica: no una teoría del conocimiento en el sentido filosófico, sino una investigación del conocer en la práctica por el cual, en nuestra sociedad, tratamos de resolver las controversias en torno a políticas a falta de una base generalmente aceptada para resolverlas¹.

Empleamos el término *discurso de políticas* para referirnos a las interacciones de individuos, grupos de interés, movimientos sociales e instituciones por medio de los cuales situaciones problemáticas se convierten en problemas de políticas, se fijan agendas, se toman decisiones y se emprenden acciones². Reconocemos el análisis de políticas como una forma de actividad intelectual que puede funcionar como causa o consecuencia de movimientos que ocurren dentro de los procesos de un discurso de políticas más general. En particular, nos interesan las condiciones en que el discurso de las políticas puede llegar a reflejar los marcos. Analizaremos estas cuestiones en el dominio específico de las políticas de bienestar.

Enmarque

La idea de los marcos ha recibido, recientemente, buena difusión. Minsky (1978) introdujo el término *marco* en el campo de la inteligencia artificial para referirse, por ejemplo, a una forma particular de representar el conocimiento. Kahneman y Tversky (1974) escribieron acerca de los marcos que distorsionan nuestra interpretación y análisis de los datos, tomando como punto de referencia una concepción particular de la toma de decisiones racional. El sociólogo William Gamson (Gamson y Lasch, de próxima aparición) ha empleado el término *empacar* para referirse a un

tipo particular de enmarque, a saber: el proceso por el cual encarnamos una idea (o marco) central organizadora en una posición de política que luego expresamos mediante símbolos tan condensadores como metáforas o lemas. Goffman (1974) ha desarrollado una especie de *análisis de marcos*, que sirve básicamente para explicar las estructuras que dan forma a los procesos de interacción social y comunicación.

Nuestro empleo de la idea de *enmarque* tiene un aire de familia con todo lo anterior, pero deseamos reservar el término a un proceso más fundamental en relación con el cual esos otros usos pueden verse como variaciones especializadas. El enmarque es, según nuestro empleo del término, una manera de seleccionar, organizar, interpretar y dar sentido a una realidad compleja en tal forma que nos ofrezca puntos de guía para conocer, analizar, convencer y actuar. Un marco es una perspectiva desde la cual se puede dar sentido a una situación amorfa, problemática y mal definida, y actuar en consecuencia.

El enmarque transpira en tres niveles: la vida personal, la investigación científica o erudita y la hechura de las políticas. Además, a menos que queramos suponer un mundo de políticas, no habitado por personas y carente de conocimiento, deberemos relacionar entre sí estos tres niveles de enmarque.

Estudiosos de muy diversas disciplinas han descubierto independientemente la importancia de este sentido más fundamental del enmarque. El antropólogo Clifford Geertz (1983) reconoce que es necesario cuando observa que: «el mundo es un lugar variado, y mucho puede ganarse (...) confrontando el espléndido estado de cosas, en lugar de desear que desaparezca en un halo a base de generalidades sin fuerza y falsos consuelos».

El filósofo social Geoffrey Vickers (1975) ha acuñado el término *sistema apreciativo* para referirse a los sistemas de valores, preferencias, normas e ideas sobre la base de los cuales enmarcamos la grandiosa realidad del mundo, y con ello lo hacemos coherente ante nosotros mismos. En su opinión, un sistema apreciativo debe satisfacer tres condiciones:

En primer lugar, debe corresponder a la realidad, lo suficiente para guiar la acción. En segundo, debe ser lo bastante compartido por nuestros compañeros para mediar en la comunicación. En tercer lugar, debe ser lo bastante aceptable a nosotros mismos para hacer tolerable la vida.

Es, pues, un constructo mental, en parte subjetivo, en gran parte intersubjetivo, es decir, basado en un juicio subjetivo compartido y constantemente desafiado y confirmado por la experiencia.

Siempre debe haber sistemas apreciativos a partir de los cuales personas, eruditos o activistas políticos construyen sus marcos.

El filósofo Nelson Goodman (1978) introdujo el término *formación del mundo* para referirse a los procesos por los cuales enmarcamos y damos forma a las realidades de los mundos en que vivimos. Y, lo que tal vez sea más significativo, el sociólogo Karl Mannheim (1936: 265) hace largo tiempo ya introdujo la idea de una sociología del conocimiento como el estudio de

las diversas maneras en que los objetos se presentan al sujeto, de acuerdo con las diferencias de medios sociales (con el resultado de que) las estructuras mentales inevitablemente se forman de distintas maneras en los diversos medios sociales e históricos.

Estructuras mentales, apreciaciones, formación del mundo y enmarque son términos que captan diferentes rasgos de los procesos por los cuales la gente construye interpretaciones de situaciones problemáticas, haciéndolas coherentes desde nuestras diversas perspectivas, y dándonos marcos evaluativos dentro de los cuales podemos juzgar cómo actuar. Nadie está exento de la necesidad del enmarque. Las prácticas personal, académica y de política dependen de ello.

Controversias de políticas

Lo que hace problemático al enmarque es que conduce a diferentes visiones del mundo y crea múltiples realidades sociales. Grupos de intereses y de votantes, estudiosos trabajando en diferentes disciplinas y personas en diversos contextos de vida cotidiana tienen distintos marcos que los conducen a ver diferentes cosas, a hacer diversas interpretaciones de cómo son las cosas y a apoyar distintos cursos de acción, con respecto a lo que se debe hacer, cómo se debe hacer y por quién.

Si los hombres ven el mundo como algo diferente para cada uno y actúan de acuerdo con sus diversas opiniones, entonces el mundo mismo se vuelve diferente. Expectativas, creencias e interpretaciones moldean los mundos en que vivimos. La obra de Alfred Schutz sobre la estructura de la vida cotidiana es un enfoque filosófico al problema de las múltiples realidades construidas, mediadas a través de signos, símbolos y lenguajes, que conducen a diferentes cursos de acción social (Schutz y Luckmann, 1981). Wallace Stevens (1954: 165) expresó muy bien una idea similar en un poema breve pero incisivo:

Dijeron: «tienes una guitarra azul,
No tocas las cosas como son».
Él replicó: «Las cosas como son
Se cambian con la guitarra azul»³.

Las políticas siempre se tocan en una «guitarra azul», porque definen y hasta cierto punto crean la forma en que son las cosas. Cuando la gente no se pone de acuerdo acerca de una cuestión de políticas, puede examinar los hechos de la situación y determinar quién tiene la razón; los *desacuerdos* en torno a políticas surgen dentro de un marco común y se pueden resolver, en principio, recurriendo a reglas establecidas. Pero las *controversias* de políticas no pueden resolverse recurriendo tan solo a los hechos ni, en realidad, recurriendo a evidencia de ninguna índole, porque se derivan de marcos conflictivos, y el mismo cuerpo de evidencia puede emplearse para apoyar posturas totalmente distintas respecto de las políticas.

Los marcos de políticas y sus subyacentes sistemas apreciativos se revelan por medio de los *relatos* que los participantes están dispuestos a hacer acerca de situaciones de las políticas. Estas narraciones, que precisan los problemas y frecuentemente basadas en metáforas generativas, enlazan los relatos causales de los problemas de política con propuestas particulares de acción, y facilitan el salto normativo que va del «ser» al «debe ser» (Rein y Schon, 1977). Dado que la realidad de cualquier situación de las políticas siempre es más rica y más compleja de lo que se puede captar por medio de cualquier relato en particular, las controversias sobre políticas están sujetas, por su esencia misma, a relatos hechos desde múltiples perspectivas. Siempre tienen el potencial de interpretación en

términos de marcos múltiples y conflictivos, donde no existen esos marcos comúnmente aceptados para la resolución. Por tanto, levantan el espectro del relativismo epistemológico.

En la filosofía de la ciencia, Kuhn (1962) ha distinguido los periodos de ciencia normal (cuando los científicos actúan dentro de un paradigma compartido [marco, en nuestro sentido] y están de acuerdo en las reglas del juego para zanjar desacuerdos) de los periodos de revolución científica, en que el desacuerdo científico pasa a través de los paradigmas y no hay un marco aceptado para zanjar disputas. En el último caso, las controversias pueden desaparecer porque quienes sostenían un paradigma competitivo sufren una conversión, o porque simplemente se mueren y nadie los reemplaza.

En su obra *Philosophy and the Mirror of Nature*, Rorty (1979) ha tratado de manera más general un tema semejante. Ha distinguido entre el *discurso normal* y el *anormal* en la ciencia, así como en otros campos de investigación. Según el sentido que le da, *normal* se refiere al discurso que procede de acuerdo con un conjunto compartido de reglas, suposiciones, convenciones, criterios y creencias, todo lo cual nos dicen cómo, a lo largo del tiempo y en principio, se puede resolver un desacuerdo. Aquí, aun cuando en realidad pueda persistir un conflicto, hay una creencia —tal vez ilusoria— basada en la suposición de que las reglas ordinarias del discurso «encarnan normas aceptadas para llegar a un acuerdo». Por contraste, el *discurso anormal* ocurre en aquellas situaciones en que las normas comúnmente aceptadas para llegar a un acuerdo no son los elementos esenciales en los que se basa la comunicación entre los actores opuestos. Tales situaciones no son definidas por los participantes en función de un marco objetivo dentro del cual se pueden arbitrar o administrar los desacuerdos. Y sin embargo, aun aquí, como lo ha observado Geertz (1983: 223), comentando la obra de Rorty:

No se abandona la esperanza de llegar a un acuerdo. En ocasiones la gente cambia de opinión, o tiene sus diferencias sobre el resultado de la inteligencia concerniente a lo que creen los individuos o grupos de individuos cuya mente corre «en otra pista».

Según Rorty, la tarea consiste en «cómo estudiar el discurso anormal desde el punto de vista de algún discurso normal». Hay algo muy atractivo en esta formulación porque suena a verdad, aunque los términos *normal* y *anormal* parezcan innecesariamente peyorativos. Geertz propone el uso del discurso *estándar* y *no estándar*, pero esa terminología parece tan insatisfactoria como la sugerencia de Rorty de que se emplee el término *hermenéutica* para describir el discurso anormal, y el de *epistemología* para describir el discurso normal. Ninguna de esas formulaciones parece correcta; ninguna parece ofrecer una alternativa clara a una visión relativista de los conflictos de enmarque.

La interpretación relativista extrema de los marcos lleva a sostener que todos los marcos son igualmente válidos. Pero aunque sea difícil refutar la lógica relativista a la que conduce un reconocimiento de enmarque, esa lógica ofende el sentido común y el entendimiento común. No todos los marcos ni todos los relatos en que se los expresa son igualmente aceptables o convincentes. Al parecer, tenemos unas normas implícitas y tal vez hasta consensuales con las cuales podemos juzgar lo adecuado de los diferentes marcos para la interpretación, el entendimiento y la acción.

Por ejemplo, March ha sugerido que nos dejemos guiar en nuestro juicio de diferentes marcos por las normas de belleza, verdad y justicia (March y Olson, 1975). La *belleza* se refiere a la elocuencia que haya en la formulación de un argumento, especialmente a la parsimonia en sus cadenas de inferencia. La *verdad* tiene que ver con la verificabilidad y confiabilidad de las implicaciones de las premisas contenidas en el argumento asociado con un marco. La *justicia* se deja guiar por normas para las proposiciones normativas contenidas en el enmarque de una situación, y en la resultante opinión de lo que es correcto o incorrecto hacer.

En muchos aspectos, estas tres normas de evaluación de marcos pueden entrar en conflicto mutuo. Sin embargo, intuitivamente comprendemos que no todos los marcos son de igual valor. Podemos discernir un terreno que se encuentra en algún lugar entre el positivismo extremo, según el cual todas las controversias importantes pueden resolverse por referencia a los hechos y la lógica, y el relativismo extremo, según el cual un marco de una situación es tan bueno como cualquier otro⁴. Pero las normas propuestas

por March para la adecuación de los enmarques no disipan el espectro del relativismo ni nos enseñan «cómo estudiar el discurso anormal desde el punto de vista de algún discurso normal». Por ejemplo, pueden resurgir conflictos de marcos cuando aplicamos las normas de la verdad, la belleza y la justicia a la evaluación de los marcos que intervienen en una controversia de políticas específica.

Un estudio del discurso de los marcos en la política social puede revelar los modos en que en realidad actuamos en nuestra sociedad, con las situaciones epistemológicas planteadas por un conflicto de marcos. Una epistemología empírica puede ayudar a aclarar cuáles normas aplicamos en realidad en nuestro juicio sobre lo adecuado de los marcos, y por medio de qué procesos enfocamos los conflictos de marcos a falta de un marco comúnmente aceptado para resolverlos. Ese estudio se enfrentaría a la cuestión de *los cambios de marco*: cómo los marcos que plantean problemas en materia de política social cambian con el paso del tiempo. Ayudaría a aclarar las propiedades de un posible *discurso que reflejara el marco*: un discurso de políticas en que los participantes reflejaran los conflictos de marcos que están implícitos en sus controversias, y exploraran los potenciales para su resolución.

Una epistemología empírica también enfocaría las posibles funciones de un tipo particular de análisis de políticas. El análisis convencional de políticas se preocupa por la elección; pregunta cómo el que toma las decisiones puede elegir racionalmente entre diversas opciones de política para realizar sus valores. Por lo contrario, el análisis de políticas que critica los marcos intenta poner en relieve el discurso de políticas que refleja los marcos, identificando las suposiciones que se dan por sentadas y que subyacen en nuestros entendimientos y acciones —aparentemente naturales— en una situación problemática de políticas. Trata de explicar los marcos conflictivos inherentes a las controversias de políticas, de modo que podamos reflexionar sobre ellos y captar mejor las relaciones que hay entre las premisas ocultas y las conclusiones normativas.

Pero el estudio de los marcos y de los conflictos de marcos se ve obstaculizado, lamentablemente, por dificultades conceptuales y prácticas, casi tan difíciles de identificar como de superar. Si queremos avanzar

hacia un discurso de políticas que refleje los marcos, antes deberemos comprender qué lo hace tan difícil.

Dificultades al estudiar los marcos

Aunque los marcos ejercen una poderosa influencia sobre lo que vemos y descuidemos, y sobre cómo interpretamos lo que vemos, paradójicamente son difíciles de evaluar. Como son parte del mundo natural —que damos por sentado—, a menudo, no nos damos cuenta de su papel al organizar nuestras percepciones, pensamientos y acciones.

En la práctica, es muy difícil distinguir entre los desacuerdos que ocurren dentro de un marco y los conflictos y controversias que ocurren entre diversos marcos. Esto se debe, en parte, a que los marcos son acerca de acción, y el deseo de hacer algo suele conducir a un compromiso de volver realizable esa acción que buscamos. A menudo lo hacemos «aferrándonos» a un marco dominante con sus metáforas convencionales, con la esperanza de «comprar» la legitimidad para un curso de acción inspirado, en realidad, por diferentes intenciones. De ahí la común discrepancia entre lo que decimos y lo que en realidad queremos decir. Como resultado, con frecuencia somos incapaces de decir cuál marco está en realidad subyacente en una posición de políticas. Cuando los participantes en un discurso de políticas empiezan a «apostar» oscurecen los marcos subyacentes.

El discurso de marcos es difícil de estudiar, por varias otras razones. El mismo curso de acción puede ser congruente con marcos totalmente distintos; por ejemplo, en las políticas de bienestar norteamericanas, hubo una marcada continuidad durante los gobiernos de Ford y de Carter, aun cuando los portavoces de sus políticas hubiesen adoptado muy distintas posiciones⁵. El mismo marco puede conducir a diferentes cursos de acción; los liberales que sostienen muy similares posiciones de política, sin embargo, están en desacuerdo mutuo acerca del trato debido a los que son ineludibles para los pagos de asistencia social.

Los marcos hacen referencia a sí mismos, pero no se interpretan a sí mismos. Una discusión acerca de los marcos casi siempre es sobre un aspecto profundo de nosotros mismos. Por tanto, no es posible separar

el marco, o sea el objeto acerca del cual queremos teorizar, de la persona que lo sostiene. Pero, como acabamos de observar, un marco no determina una posición particular acerca de una cuestión sustantiva de política, y muchas posiciones de políticas pueden ser congruentes con un marco determinado. Hay más que una conexión lógica entre un marco y sus consecuencias prácticas para la acción.

Las políticas, en términos generales, tienden a no evolucionar por medio de cambios de enmarque autorreflexivos, de alcance general, del tipo que Heclo (1985) ha descrito como «cambios en la autoconciencia»: «Los cambios en la autoconciencia implican los más generales cambios de paradigma [marco]: ¿Dónde estamos? ¿Por qué estamos aquí? ¿Adónde vamos?». Por otra parte, tenemos muchos ejemplos de cambio de marco en torno a cuestiones específicas de política. Por consiguiente, necesitamos enfocar el enmarque de las cuestiones de políticas que pueden estar o no estar relacionadas con cambios de marco más generales. Necesitamos preguntar qué es lo problemático en estas cuestiones, en qué terrenos de políticas se encuentran y cuáles factores explican las formas en que están enmarcados. Sabemos mucho menos acerca de estas cosas de lo que solemos creer... aunque, sin duda alguna, entran en juego intereses materiales y simbólicos.

Es difícil distinguir entre los cambios de enmarque reales y los potenciales. Por ejemplo, en el campo de la política de asistencia pública, la introducción del ingreso complementario del seguro (SSI) podría interpretarse como un reenmarque de las políticas norteamericanas hacia los pobres, porque significó nacionalizar las instituciones y la legislación que solemos relacionar con la Ley para los Pobres, integrándolas en el marco de la seguridad social, al menos para una parte importante de la población que depende de la asistencia. Este acontecimiento pudo haber creado un potencial para un cambio de enmarque en algún momento posterior; pudo conducir a reenmarcar el problema de la pobreza en términos de una concepción más general de la responsabilidad social de ayudar a los pobres. Pero ese potencial de reenmarque no se realizó. No se introdujeron otras reformas, por lo cual el potencial de un reenmarque más vasto quedó latente. En realidad, lo que ha ocurrido es que los ancianos parecen estar disfrutando de una posición privilegiada a expensas de las familias pobres y los niños dependientes.

Es posible lograr reenmarque sin controversias y controversias sin reenmarque. Las políticas pueden cambiar sin la elección de un marco, y podemos discutir sobre la elección de marco sin ningún cambio político. Por ejemplo, durante los sesenta y comienzos de los setenta, surgió en los Estados Unidos un debate por la introducción de un impuesto negativo general sobre la renta. Se efectuaron muchos experimentos sociales para mostrar que el ingreso garantizado no perturbaría el esfuerzo laboral ni la estabilidad de las familias, y se introdujeron reformas legislativas. Pero el plan abortó; de ahí no surgió ningún gran cambio. Mientras tanto, las políticas públicas norteamericanas seguían comprometidas con el concepto de que las personas debieran tener un ingreso puro en efectivo, en lugar de prestaciones en especie o asignadas con propósitos específicos, como vivienda, atención médica o alimentos. Pero en la práctica, pese al compromiso con la expansión de las opciones individuales, lo que surgió fue una creciente proporción del gasto para beneficios en especie, que limitó la elección individual a ciertas categorías particulares de gastos como combustible, alimentos y atención médica. Hubo así un importante reenmarque de las políticas, pero sin una explícita decisión pública de avanzar en dirección del reenmarque.

Muchos cambios de política consisten tan solo en adaptarse a situaciones cambiantes. Empero, el efecto acumulativo de muchas de estas adaptaciones puede *entrañar* un reenmarque del modo en que pensamos y actuamos con respecto a la cuestión de políticas de que se trate. Podemos vernos llevados a un cambio de marcos por algo distinto de la evolución de las ideas o la experiencia de la investigación dentro de un marco.

Los diversos tipos de fenómenos que obstaculizan una epistemología empírica del discurso de políticas pueden ser, en sí mismos, objeto de estudio. Un estudio de los conflictos de marcos y los cambios de enmarque debe enfocar el enmarque de cuestiones particulares de política y aguardar a la interrelación de procesos tales como el «aferrarse», enmarcar al nivel de las políticas adoptadas y aplicadas, y la obligación de reenmarcar por medio de múltiples adaptaciones a circunstancias cambiantes.

Un vocabulario preliminar para el estudio de los marcos

Nombrar y enmarcar

Una vez nombrado un terreno de políticas, su nombre parece natural. Pero el paso de una sensación incoherente de que algo, en una situación, está mal, a una especificación de lo que está mal puede estar preñado de incertidumbre, intelectual y emocional en sus manifestaciones.

En *The Feminine Mystique*, Betty Friedan (1964) describe la sensación que tuvo, durante los sesenta, cuando se sentía preocupada por la posición de las mujeres pero no sabía qué nombre dar a la angustia que sentía. El *sometimiento femenino* fue el nombre dado al fenómeno que le parecía tan perturbador. El nombre mismo indicó implícitamente el camino a la solución del problema, a saber: la *liberación femenina*. Desde la perspectiva del marco resultante, cuestiones políticas específicas —por ejemplo, las cuestiones de la atención durante el día y los pagos de asistencia a las mujeres— adquirieron significados muy distintos.

El nombre asignado a un terreno problemático llama nuestra atención hacia ciertos elementos y nos lleva a pasar otros por alto. Organizar las cosas nombradas las une en un todo integrado. El proceso complementario de nombrar y de enmarcar construye socialmente la situación, define cuál es su problemática y sugiere cuáles son los cursos de acción apropiados. Nos da coherencia conceptual, una dirección para la acción, una base para la persuasión y un marco para la recabación y el análisis de datos: orden, acción, retórica y análisis.

Cuando los participantes en una controversia de políticas nombran y enmarcan la situación de las políticas de maneras distintas, a menudo es difícil descubrir por qué están luchando. Por ejemplo, no es posible decir, sencillamente: «Comparemos diferentes perspectivas para resolver la pobreza», pues es probable que cada enmarque del problema de la pobreza seleccione y nombre diferentes características de la situación problemática. Ya no somos capaces de decir que estamos comparando distintas perspectivas sobre «el mismo problema», porque el problema mismo se ha modificado. Por ejemplo, aunque podamos ponernos de

acuerdo en que la pobreza trata de falta de recursos, la naturaleza de lo que falta puede ser sumamente oscura. La idea de *dependencia* hace problemático el proceso por el cual los pobres adquieren recursos, tratándola como pauperización. Las transferencias de ingreso tendientes a responder a la falta de recursos pueden crear nuevos problemas de dependencia, o una subclase que obtenga su ingreso gracias a la generosidad del gobierno. Cuando se considera la pobreza en función de la desigualdad hay una preocupación por la distribución relativa de recursos, de modo que al dar ayuda a los pobres nos vemos obligados a considerar la posición económica de los grupos de ingresos altos y medianos de la sociedad. En cada caso, el nombre dado a la problemática situación de la pobreza selecciona diferentes fenómenos —en el mejor de los casos, traslapantes— para centrar en ellos nuestra atención y los organiza de manera distinta.

El contexto

El enmarque de una cuestión de políticas siempre ocurre en el seno de un contexto anidado. Las cuestiones de políticas suelen surgir en relación con programas gubernamentales, que existen en algún entorno de políticas, el cual a su vez es parte de otro medio político y económico más vasto, localizado a su vez dentro de una época histórica. Cuando se modifica algún rasgo del contexto anidado, los participantes pueden descubrir que ya no funciona la repetición de una fórmula que antes había tenido éxito. Entonces, esa percibida modificación del contexto puede fijar el clima dentro del cual redes adversarias intentan reenmarcar una cuestión de política, dando otros nombres al terreno de las políticas, reconstruyendo las interpretaciones sobre cómo las cosas llegaron a ser lo que son y proponiendo lo que puede hacerse al respecto.

Resulta útil distinguir al menos cuatro contextos anidados. Un programa puede servir como su propio contexto *interno*, que se modifique con el tiempo mediante el remplazo de su personal, sus patrocinadores o sus clientes. Las metas surgen de las posibilidades de la situación interna y de la necesidad de adaptarse a los cambios que ocurren en ella. Característicamente, la organización se adapta a su nueva situación en lugar de intentar deliberadamente modificar su marco. Pero unas

nuevas reglas de hacer las cosas, adoptadas con espíritu de ajuste, pueden preparar los fundamentos de un cambio considerable del marco, corriente abajo.

El contexto *próximo* es el entorno de política en que opera un programa. Un programa siempre existe en interacción con otros programas, y el reenmarque puede ocurrir cuando se modifica el contexto próximo. Por ejemplo, el alojamiento público en los Estados Unidos fue redefinido en el curso del desarrollo de un programa de renovación urbana que arrojó a una diferente clase de gente a la vivienda pública, modificando así el mundo del programa y obligándolo a adaptarse a un medio nuevo.

El contexto *macro* incluye los cambios en las direcciones de las políticas, los cambios de las instituciones destinadas a aplicar medidas políticas, la realineación de la política partidista y las fluctuaciones económicas. Los cambios ocurridos en el contexto macro no necesariamente fijan las condiciones para el reenmarque de las políticas, al menos a corto plazo, pero sí pueden conducir a un reenmarque simbólico. Así, como hemos visto, la retórica del desacuerdo puede sugerir un importante reenmarque de las políticas, mientras la práctica muestra una notable continuidad.

Los cambios *globales* de contexto son más difíciles de especificar, aunque sean de extrema importancia. Incluyen cambios de las eras históricas en que puede ocurrir un reenmarque de las cuestiones de política. Poco se ha escrito acerca de las eras históricas de tal forma que vuelva a vincularlas con las políticas públicas, los entendimientos científicos y las percepciones individuales; sin embargo, un cambio de una era histórica puede ejercer un notable impacto sobre el enmarque de las cuestiones de política. Por ejemplo, al parecer acabamos de dejar atrás la era de los sesenta con su evidente compromiso con la innovación institucional en favor de los pobres y los desprotegidos. La era de los ochenta aún parece estar desarrollándose, pero temas como la reprivatización y los límites de la intervención gubernamental parecen evidentes, acaso como reacción al afán expansionista de la época anterior.

Hace mucho tiempo, Innis (1951) nos recordó el célebre aforismo de Hegel de que «el búho de Minerva levanta el vuelo al caer la noche», es decir,

la sabiduría de una época solo llega cuando está terminando. Esta es una observación cínica, porque es precisamente en el punto en que una era ha terminado cuando ya no podemos hacer nada por ella. Sin embargo, tal vez Hegel tuviera razón, en parte. Quizá un cambio de contexto nos muestre que algo se ha modificado en una sociedad, aunque el cambio aún no haya sido plenamente captado. Un cambio de contexto puede verse como el instintivo tanteo de una sociedad que trata de comprender, de antemano, el pleno desarrollo del cambio social. Por tanto, el búho de Minerva puede volar al alba mejor que al anochecer. Las ideas y los símbolos del sistema cultural de la sociedad (en términos de Parsons) pueden prefigurar los cambios de su sistema social antes de que estos cambios hayan sido cabalmente comprendidos (Parsons, 1967).

Hasta aquí hemos dicho que el enmarque de las cuestiones de política responde a los cambios de los contextos más vastos de la política, cambios que a menudo entrañan alteraciones del propio marco. Pero también es cierto que el reenmarque de cuestiones puede dar forma a los contextos de los cuales depende. Algunos cambios de contexto ocurren cuando individuos y organizaciones se adaptan a sus situaciones locales; el efecto acumulativo de sus adaptaciones produce un nuevo mundo social que, a su vez, conduce a nuevas formas de conducta. El reenmarque de cuestiones de políticas brota de los cambios de contexto y también ayuda a producirlos. Tanto la adaptación como el aprendizaje social son operativos.

El discurso del marco

La conversación interpersonal tal vez sea el sentido fundamental del término discurso. El empleo de este término para referirse al diálogo dentro de las instituciones y entre ellas es una extensión metafórica. Sin embargo, el caso prototípico de discurso interpersonal puede interpretarse al menos de dos maneras distintas. Podemos ver a dos personas hablando entre sí acerca de un asunto en que ambas están directamente interrelacionadas. Por ejemplo, cada una de ellas desea algo que solo una puede tener. En este caso, la *controversia de política* es de naturaleza interpersonal. O puede verse a dos personas hablando entre sí acerca de una cuestión de políticas más general, en relación con la cual funcionan como agentes de grupos o

instituciones que son partes de la controversia de políticas. En este caso, la controversia es de naturaleza institucional.

En cualquiera de estos casos, mientras el discurso ocurra en un contexto interpersonal será importante hacer referencia al *mundo conductual* en que está ocurriendo. Cuando las personas hablan entre sí sus acciones tienen significados múltiples: no solo comunican mensajes cuyos significados pueden ser contruidos por el otro, sino también transmiten actitudes de segundo nivel hacia la interacción misma. También modelan los modos de conducta que la otra parte podrá aceptar. Presentan teorías en uso para la interacción interpersonal (Argyris y Schon, 1974). Así, como lo ha señalado Bateson (1972), el discurso interpersonal puede adoptar la propiedad de la esquizogénesis, o «cuanto más, más». Por ejemplo, cuanto más un bando defiende su posición y trate de ganar, más lo hará también la otra parte. Las pautas de interacción interpersonal pueden contribuir a la creación de un mundo conductual que esté más o menos abierto o cerrado, sea desconfiado o confiado, auténtico o engañoso, defensivo o no defensivo, contencioso o cooperativo. Cada jugada en una interacción interpersonal tiene significado para el mundo conductual, así como para la cuestión de políticas, que los participantes tratan como tema de su discurso.

El discurso interpersonal también debe tener un lugar institucional dentro de algún sistema social más general. No hay vacíos institucionales. Hasta una charla entre amigos íntimos ocurre en el medio institucional de «la casa de alguien» o «un paseo por el parque». Cuando los eruditos hablan entre sí acerca de una controversia de políticas, lo hacen dentro del contexto institucional de una universidad o de un instituto de investigación. Este nicho institucional es importante por varias razones para la naturaleza del discurso. El contexto institucional puede llevar sus propias perspectivas características y modos de enmarcar los asuntos, como lo hemos notado antes; o bien puede ofrecer particulares roles, canales y normas para la discusión y el debate; el discurso tiende a adaptarse a las normas de las instituciones en que se encuentra incrustado. Por ejemplo, en un tribunal en que se espera una interacción entre adversarios, cada quien tiende a suprimir sus dudas acerca de su propia posición. Ante la mesa de negociaciones, cada locución tiende a interpretarse como una jugada en un juego de regateo. Sin embargo, aun en medios como estos, las personas

pueden emplear su libre albedrío para actuar en forma desviada, violando así las normas institucionales.

Cuando el discurso es *público*, adopta las formas especiales de las instituciones reservadas en nuestra sociedad a las interacciones públicas acerca de controversias de políticas de preocupación pública: los *foros de políticas*, que sirven de vehículos institucionales para el debate de políticas. Incluyen las arenas legislativas, los tribunales, las comisiones públicas, los consejos de gobierno y de partidos políticos, las páginas editoriales de revistas y periódicos, los programas de radio y televisión, y las salas de seminario y de conferencias de las academias.

Los foros de políticas tienen sus propias reglas del juego. Ahí, por una parte, las locuciones individuales probablemente tendrán significados y consecuencias alejados del contexto interpersonal en que ocurren. Por ejemplo, podemos hablarnos unos a otros, pero nuestras palabras posiblemente sean registradas y repetidas ante un público más numeroso, apartadas de su inicial interacción cara a cara, y en ese contexto más general tal vez no sepamos cómo los otros interpretarán y responderán a nuestras palabras. El discurso puede estar enmarcado dentro de la arena de la política formal, interpretado como si favoreciera a uno u otro bando en una pugna continua entre partidos políticos. Puede ser modificado por los medios de la comunicación pública: la prensa, las publicaciones especializadas, la radio, la televisión y similares.

Entre las reglas del juego en los foros de política están las normas por las cuales se hacen juicios acerca de la legitimidad de los participantes: su categoría como participantes en la conversación de políticas. Las disputas dentro de un discurso de políticas pueden tener que ver con la lucha por obtener categoría. Los que fueron excluidos de la conversación pueden luchar por intervenir. Como un aspecto de este tipo de lucha, ciertos movimientos sociales pueden entrar en acción para poner ciertas ideas en buena circulación y legitimar a sus propios representantes como participantes en una conversación de políticas, empleando con este fin todo el repertorio de la acción y la protesta sociales. De este modo, por ejemplo, el movimiento por los Derechos de Asistencia, del decenio de 1960, intentó modificar los términos de referencia del debate por la asistencia social,

e introducir participantes nuevos como actores legítimos en los foros de políticas de la época.

En los foros de políticas y a su alrededor hay toda una gama de roles sociales que ocupan diferentes posiciones en el proceso del discurso de las políticas. Por ejemplo, hay una fundamental división del trabajo entre los interesados por la movilización de la acción y los preocupados por la movilización del consenso intelectual. Sin embargo, como en la mayoría de tales situaciones, el límite suele ser borroso. Mediar entre el pensamiento y la acción lleva a algunos a subrayar el primero y a otros a preferir la segunda, mientras que otros más tratan de tender un puente entre uno y otra, combinando la investigación y la práctica⁶.

Como ya hemos observado, los marcos nunca son autointerpretativos. La interpretación de asuntos particulares de políticas en términos de varios marcos siempre la emprende alguien: habitualmente grupos de individuos, o bien organizaciones formales o informales. Estas entidades actúan como patrocinadoras del marco. Cuanto más naturales sean los marcos que moldean nuestro pensamiento y cuanto más los demos por sentados, menos probablemente cobraremos conciencia de nuestros marcos y a la vez de los transmisores sociales que interpretan las cuestiones de políticas en términos de marcos.

Los patrocinadores de un marco tratan de desarrollarlo, de explicar sus implicaciones para la acción y de echar las bases para los argumentos. También podrán inventar metáforas para la comunicación acerca del marco: metáforas diversamente relacionadas con las que acaso hayan contribuido a la generación del marco mismo.

El patrocinio de marcos puede correr a cargo de organizaciones de investigación dentro de la comunidad de las ciencias sociales. Aquí, los analistas de las políticas pueden desempeñar una función crítica en la creación de marcos, trabajando dentro y fuera de las burocracias gubernamentales en papeles de empresarios, intermediarios y corredores de ideas. Ellos pueden nombrar el terreno de las políticas y especificar cómo deberán vincularse los marcos, los diseños y las acciones de las políticas. Pueden actuar como especialistas técnicos que depuran los

problemas que surgen en el enmarque de una cuestión de política y, en el proceso, ponerla en buena circulación⁷. Pueden combinar investigación y experiencia en el uso de símbolos, metáforas comunicativas y suposiciones simplificadoras. De todas estas maneras, ayudan a crear el marco; pero estos procesos mismos pueden llevarlos a los límites de sus marcos y, por tanto, al reenmarque.

Los *intelectuales de las políticas* desempeñan el papel especial de explicar las ideas y marcos de políticas implícitos en la acción de los movimientos sociales. Por ejemplo, en este sentido, Piven y Cloward (1971) se encontraron entre los intelectuales de las políticas que ayudaron a sacar a la superficie y a defender las ideas implícitas en las acciones del movimiento por los Derechos de Asistencia.

Los tres niveles del discurso de políticas —el individual, el interpersonal y el público— están interrelacionados en formas complejas. Cada nivel superior de agregación presupone la existencia de los que están debajo. En el discurso interpersonal suponemos que los individuos también llevan adelante un diálogo paralelo en sus propias mentes, pensando para sí mismos mientras hablan a otros. En el discurso público hay modos de diálogo interpersonal. Los representantes individuales de las instituciones en pugna pueden reunirse para una discusión informal y hablar entre sí ante la mesa de negociaciones, aun cuando sus interacciones tendrán allí una mayor resonancia institucional.

A la inversa, cada nivel inferior de discurso puede reflejar el nivel superior en relación con el cual ocurre. Las personas pueden hablar entre sí de un modo que muestre que están tomando en cuenta a los partidos políticos o a los diversos grupos de votantes a los que pertenecen. Y aun en la intimidad de su fuero interno, pueden emplear el lenguaje, obedecer las normas y experimentar los sentimientos normales en la conducta de una controversia sobre política pública.

El interjuego de los niveles del discurso, la manipulación de los contextos institucionales del discurso, puede desempeñar partes importantes en la investigación que da forma a la carrera de una controversia de políticas. Por ejemplo, el problema de lograr una resolución satisfactoria de una

disputa de políticas puede estar enmarcado de tal manera que permita a los individuos *liberarse* de las restricciones del debate de las políticas públicas. Por ejemplo, la gente dice: «juntemos a los principales actores donde puedan interactuar como seres humanos» (llevándolos «a lo alto de la montaña», donde pueden entrar en «su espíritu»). O bien, el problema de resolver una controversia puede enmarcarse en tal forma que mejore el funcionamiento de las instituciones públicas en cuestión: por ejemplo, mejorando la estructura de un proceso de negociación colectiva. Si dos personas tienen una relación especial que les da una sólida base de confianza mutua y para una indagación cooperativa, aunque estén sentadas «en lados opuestos de la mesa», su relación especial podrá servir como vehículo para resolver la controversia. Si una persona puede encontrar en sí misma toda una multiplicidad de opiniones diferentes y conflictivas sobre el asunto en cuestión, puede ser que llegue a estar más dispuesta a aceptar las opiniones de los adversarios con quienes haya trabado la disputa.

Cambios de marco: el ejemplo de las políticas de bienestar

En la compleja interacción de quienes participan en un discurso de marco a través de múltiples tipos y niveles de contexto, ¿cómo ocurren en realidad los cambios de marco? Que nosotros sepamos, hay pocos o ningún estudio empírico del reenmarque. Lo que sigue es un relato, un tanto especulativo e interpretativo, de uno de los modos en que ocurren los cambios de enmarque en el área especial de las políticas de bienestar⁸.

Desde 1962 ha habido al menos tres marcos dominantes en el terreno de la política de bienestar. Los llamamos *capital humano*, *incentivos* y *obediencia a las normas*.

Una de las suposiciones críticas que sirvieron de premisa a la Guerra contra la Pobreza fue captada por la imagen: «Una mano (...) no un puñado». Se haría el hincapié programático en los servicios, especialmente en el entrenamiento y la preparación de la mano de obra. Se pensó que los pobres y los que recibían pagos de asistencia pública necesitaban adquirir capacidades que les permitieran participar en el mundo laboral. Se dio apoyo a los programas de creación de empleos cuando la experiencia

indicó que no bastaba la preparación institucional en las aulas. Estos empleos tenían ambientes protegidos en que, se pensaba, las personas lograrían adquirir habilidades relacionadas con el trabajo.

Un tanto paradójicamente, la siguiente etapa de la evolución del enmarque de las políticas de asistencia pública subrayó el puñado, antes que la mano. Después de todo, la preparación requería largo tiempo y no se podía dar a todos. La gente necesitaba más que capital humano; también necesitaba dinero.

Los patrocinadores del enfoque de los *incentivos* creyeron que la pobreza trataba básicamente de falta de dinero, y de un sistema que consideraban esencial para redistribuir el ingreso a quienes se encontraran fuera del mercado laboral. Estas personas inventaron —mejor dicho, reinventaron— un mecanismo para redistribuir el ingreso, vinculado con el sistema del impuesto sobre la renta popularmente conocido como impuesto negativo sobre la renta. Los experimentos con el impuesto negativo sobre la renta pretendieron mostrar que redistribuir el ingreso de esa manera no afectaría negativamente la conducta en el trabajo ni perturbaría la vida de familia. En la política pública norteamericana, estas ideas se discutieron durante cerca de diez años. En las políticas de bienestar se creó un sistema de incentivos, con objeto de alentar a trabajar a quienes hacían trampas para recibir pagos de asistencia pública y, con el tiempo, a ganar el dinero suficiente para no tener que recibir esos pagos.

Con el tiempo, el esfuerzo por reenmarcar el sistema de transferencia del ingreso abortó políticamente y la idea languideció. Pero al mismo tiempo, también el contexto cambió a la época que nosotros conocemos hoy como los ochenta. Varios patrocinadores conservadores promovieron nuevos marcos para hacer frente a la pobreza.

Una primera estrategia consistiría en mostrar que el problema de la pobreza era mucho más pequeño de lo que se había imaginado. Los partidarios de este enmarque arguyeron que en el atiborrado medio de las políticas de los sesenta y los setenta se habían desarrollado muchos programas para dar a los pobres recursos económicos que no fueran en forma de pagos en efectivo. Por ello, era un error definir la pobreza tan solo como falta de

dinero. Lo que se necesitaba era un esquema más vasto, en que el valor monetario de estos servicios en especie pudiera contabilizarse y añadirlo al ingreso de los pobres. Según esta medida, podría considerarse que la pobreza se había exagerado como problema. De hecho, solo 4 o 5 % de la población era pobre, dependiendo de los servicios que se incluyeran en la definición general y del método que se empleara para contabilizarlos. Estos patrocinadores creyeron que no debiera emplearse un sistema de redistribución como sistema de incentivos para reducir la dependencia económica. Antes bien, sostuvieron, el mantenimiento del ingreso para los pobres debiera ser un programa residual, enfocado a quienes estuviesen fuera del mercado laboral. El programa general fue rechazado en favor de tratar el mantenimiento del ingreso como una «red de seguridad».

El enfoque de los incentivos fue rechazado y remplazado por un compromiso con la obediencia a las normas. Se esperó que la gente siguiera la norma de trabajar sin necesidad de incentivos para alterar su conducta. Se creyó que la política debía suprimir todos los incentivos, y se insistió en que las personas capacitadas para trabajar debían hacerlo. En situaciones ambiguas, como la de las madres solteras trabajadoras con hijos pequeños, sería necesario hacer una elección: o trabajo o pago de asistencia. Los inelégibles que recibieran prestaciones debían salir de las listas de pago.

¿Cómo podemos comprender la trayectoria de estos cambios de marco? No solo fueron resultado de la acumulación de conocimiento científico ni de cambios del poder político. No fueron resultado de que comprendiéramos mejor las cosas o supiéramos más, porque el conocimiento proveído por las ciencias sociales se utilizaba, en su mayor parte, después de que ya se había alcanzado una postura con respecto a una política y se necesitaban «municiones» para legitimarla. La gente solía primero hacer las paces en cuanto a sus opiniones y rara vez hacía uso del conocimiento de las ciencias sociales para llegar a una posición ante las políticas. Dado que los intereses no son autodefinitorios, estos cambios de marco no fueron resultado de poder o intereses; no hay una manera mecánica de saber cómo el poder calculará sus intereses. El gobierno conservador del presidente Nixon patrocinó un sistema de ingreso garantizado y de incentivos en materia de asistencia pública (el Plan de Asistencia Familiar) y otro gobierno conservador, el del presidente

Reagan, rechazó el enfoque de los incentivos y en cambio favoreció un compromiso más enérgico con la obediencia a las normas.

Pero no hay que excluir por completo el conocimiento y los intereses. Podemos interpretar este proceso de sucesión de marcos como resultado de tres procesos diferentes y paralelos del discurso de políticas, que ocurrieron en distintos contextos globales. No comprendemos muy bien todos estos procesos ni podemos articular todas sus relaciones entre sí.

Empecemos por el proceso que más a menudo se ha tratado en la bibliografía sobre la elaboración de políticas: el conflicto de los marcos de patrocinio y el problema del consenso. Según esta opinión, los marcos se asignan a diferentes patrocinadores oficiales; no entran los marcos de quienes los refutan. Por ejemplo, los radicales suelen ver la asistencia pública como un «ejército de reserva», que será útil para domar las exigencias salariales de los sindicatos en los mercados laborales activos. El ala liberal del Partido Demócrata favorece los incentivos y los enfoques del capital humano (la tensión entre estos dos enfoques salió a la superficie durante el gobierno de Carter, el cual había esperado combinarlos en un mejorado Programa de Empleos e Ingresos). Los conservadores tradicionales desean un gobierno pequeño y no creen que el gobierno deba tratar de modificar la conducta de los ciudadanos. A su parecer, deben hacerse todos los esfuerzos necesarios para mantener bajas las nóminas, y el bienestar social debe ser un programa tan reducido como sea políticamente posible. Al mostrar que la pobreza es un problema pequeño y que los programas contra la pobreza ayudan a crear las condiciones que ellos mismos trataban de mejorar, los conservadores tradicionales intentan crear un clima intelectual y simbólico en que pueda florecer semejante programa minimalista. Por contraste, los conservadores cívicos favorecen el gobierno grande y por ello están dispuestos a apoyar una garantía al ingreso, pero en cambio objetan al gobierno tolerante. Insisten en mantener las obligaciones de trabajar y en imponer pruebas de trabajo, en evitar depender de incentivos y de enfoques de capital humano. La opinión pública no aceptará las opiniones conservadoras liberales o tradicionales, y los conservadores cívicos carecen de poder político para imponer sus ideas. El resultado es un estancamiento (Mead, 1985).

Los pasajes anteriores son un relato sobre marcos en conflicto, apoyados por patrocinadores oficiales con valores en conflicto y acerca de una opinión pública ambivalente e incapaz de apoyar un marco o el otro. Es un contexto «vicario» entre patrocinadores de igual poder. Y dado que todos ellos pueden apelar a normas que resuenan con los tradicionales valores norteamericanos, los llamados simbólicos no pueden resolver la controversia.

Pero interpretar la sucesión de marcos como un choque de marcos apoyados en los foros del patrocinio oficial y divorciada del foro de la opinión pública, y por ello atrapada en un intratable dilema de políticas, sería oscurecer el nivel real del consenso que hay en el trasfondo y que traza una línea demasiado clara entre controversias irreconciliables a través de marcos y desacuerdos dentro de un marco. Otro modo de interpretar esta historia de la asociación de marcos es considerar el proceso como la carrera de un solo marco. Según esta opinión, el marco común se deriva de la idea de que la asistencia pública debe ser autolimitadora y residual. Y se considera que el principal objetivo a largo plazo de las políticas de bienestar es una reducción de las listas de asistencia pública y no un salario a los ciudadanos, basado en un derecho. Se considera que cada cambio de políticas es resultado de las insuficiencias de las políticas de la fase anterior. El carácter esencial y los efectos de cada ronda de ideas sobre políticas se revelan conforme las personas empiezan a poner en práctica esas ideas.

Una tercera interpretación de la historia considera la evolución de los marcos, en su mayor parte, en función del cambio de contexto, especialmente del económico. Durante los sesenta, los programas de incentivos y capital humano fueron aceptables porque la economía había generado un excedente, y los intelectuales de las políticas desarrollaron la idea de que si no se gastaba ese excedente, se socavaría el crecimiento económico. Al llegar los ochenta habían cambiado el contexto económico y los marcos para interpretarlo. Las cuestiones de política llegaron a ponerse en términos de una permuta (*trade off*) o dilema, que incluía dos objetivos en conflicto: promover la equidad social y fomentar la eficiencia económica. El argumento en favor de la reforma al Estado benefactor dice que la eficiencia no puede ser el objetivo supremo; también se deben tomar en cuenta cuestiones de igualdad, equidad y reducción de la

pobreza. Por contraste, la bibliografía sobre la crisis del Estado benefactor afirma que los objetivos de equidad se han llevado demasiado lejos, y que hoy existe el riesgo de socavar la capacidad de la sociedad para entrar con éxito en una competencia económica global y mantener con ello los recursos necesarios para asegurar una sociedad más equitativa. Según esta idea, ya se ha llegado a los límites de la mayor expansión del Estado benefactor. Si se quiere mantener la viabilidad económica, será necesario hacer algunos recortes.

Ante tres interpretaciones parciales de la historia, cada una de las cuales es a la vez válida e incompleta, evidentemente hemos de tratar de integrarlas. Pero aquí completamos un círculo. Se necesita un marco para desarrollar y para juzgar un marco: un modo de construir el mundo social que nos permita asimilar e interpretar los hechos de que disponemos. Consideremos el contexto: ¿estamos experimentando en realidad una crisis del Estado benefactor?

En la Organización de Cooperación y Desarrollo Económico (OCDE), importante foro internacional para el discurso de políticas, el concepto de la crisis del Estado benefactor pasó a ser tema de gran importancia a finales de los setenta y comienzos de los ochenta (OCDE, 1985). Desde luego, la propia organización depende, para su apoyo económico, de los estados miembros, entre ellos los Estados Unidos. Como organización internacional, es especialmente sensible a los ministerios financieros de diferentes países y a los políticos y burócratas, partidarios de los recortes, tanto en los Estados Unidos como en otras grandes naciones industriales. En la atmósfera de finales de los setenta, estas influencias convergieron ayudando a generar un ambiente de sentimiento, que tal vez deba caracterizarse como de crisis de la política social. Según estas ideas, el Estado benefactor era el principal causante, o al menos uno de los principales, de la decadencia de las economías de Europa y de los Estados Unidos, porque afectaba perversamente los incentivos para trabajar e impedía el gasto del Estado, inhibiendo así las inversiones y contribuyendo indirectamente a lo que algunos críticos han descrito como una huelga de inversiones por el capital. El capital estaba reservándose para encontrar mejores condiciones, así como lo hace la mano de obra cuando se declara en huelga. La mejor condición que estaba buscando era una carga reducida

al pagar por la protección social que los grupos laborales y cívicos habían conquistado a lo largo de los años.

Para cuando Reagan fue elegido presidente en 1980, la OCDE ya había patrocinado su primera conferencia internacional: «La crisis del Estado benefactor», anticipándose con presciencia a la era conservadora que inaugurarían los años de Reagan. Algunas personas han conjeturado, además, que la cultura institucional de la burocracia de la OCDE, al prever lo que podía ser una aceptable posición norteamericana, se mostró especialmente favorable a la campaña de reelección de Reagan en 1984. El discurso internacional acerca del desarrollo del Estado benefactor vendría a desarrollar más aún el marco de crisis que se había introducido por vía de la publicación de los resultados de la conferencia en 1981 (OCDE, 1981).

Aunque respondiendo a la ideología política de Reagan, la OCDE adoptó una posición un tanto distinta y más moderada a comienzos de los ochenta acerca de la contribución del Estado benefactor hacia los graves problemas económicos que experimentaban los países europeos. Podemos llegar, incluso, a decir que la OCDE cambió de opinión. ¿Por qué? Pensamos en esos términos porque el personal responsable de escribir el informe adoptó un marco diferente. El cambio de personal, más que el cambio de los hechos, condujo a una «lectura» distinta de lo que estaba ocurriendo. El informe más reciente sostiene que se exageró el argumento de la crisis y que quedaban recursos suficientes para permitir cierto modesto desarrollo en el Estado benefactor y hacer posible una innovación tecnológica y una ulterior expansión económica. Aunque se reconoce la cuestión de la crisis, en cambio se reinterpretan los datos sobre los niveles de gastos para que sean congruentes con este nuevo argumento. Aunque el patrocinador parece haber cambiado de opinión, la idea se había difundido, y probablemente ya era demasiado tarde para invertir su creciente influencia. El reenmarque de la cuestión del bienestar debe verse en el contexto de la generalizada aceptación de este marco de crisis.

La sucesión de marcos sobre la cuestión de la política de bienestar, pasando del capital humano a los incentivos y a la obediencia de las normas, ¿deberá interpretarse como un aprendizaje social dentro del marco común de «reducir las nóminas de la asistencia» o como competencia entre distintos

marcos o como ambas cosas? Las ideas asociadas al aprendizaje social acumulativo dentro de un marco no pueden separarse de un metamarco organizador, que considera la sucesión de marcos como una investigación cooperativa especialmente científica, y no como una pugna de ideologías sin reglas para arbitrar la controversia. Por otra parte, de cuando en cuando, un diferente tipo de aprendizaje social puede verse en las «coyunturas» de un proceso dialéctico del discurso de políticas, en que los participantes luchan por controlar la transición del marco de una cuestión a su sucesor. Aquí, las inadecuaciones que se han visto ya en el marco anterior de la misma cuestión pueden entrar en el debate y la investigación que median en la lucha política entre los defensores de marcos en competencia.

Wollmann (1984) ha sostenido que los analistas de las políticas y los conceptos de investigación que emplean reflejan las coaliciones de fuerzas políticas. Cuando cambian las condiciones, los conceptos pueden desfasarse. Por ello, los investigadores acaso se vean obligados a reconocer lenta y dolorosamente que los conceptos científicos que habían estado promoviendo (por ejemplo, la «evaluación de programas») sencillamente reflejaban la sabiduría tradicional aceptada de los contextos institucionales en que estaban operando. Pragmáticamente, muchos científicos sociales han comenzado a modificar sus conceptos y métodos de investigación para que reflejen el nuevo contexto en que están laborando. Se están adaptando a una situación modificada y alterando su propia conducta, lo mejor que pueden. Por otra parte, otros especialistas están experimentando con una apreciación diferente no solo de sí mismos y de los cambios que hay en la situación externa, sino también de ellos mismos *en* esa situación. Su conciencia de sí mismos es uno de los factores que han generado interés en el papel del enmarque en el discurso de las políticas.

Lo que hasta aquí ha faltado a nuestro análisis es poner atención a los usos posibles de la autoconciencia de quienes participan en el discurso de las políticas. Si los formuladores de las políticas, los políticos y los intelectuales de las políticas cobraran mayor conciencia de sus marcos subyacentes y del conflicto de marcos que ellos transmiten, si se movilizaran y actuaran basándose en una conciencia de sí mismos en la situación de las políticas, ¿cuáles podrían ser las consecuencias para el discurso del marco? En pocas palabras, ¿cuáles son las posibilidades para un discurso de políticas que

refleje el marco y para un análisis de las políticas que critique el marco dentro de semejante discurso?

El discurso que refleja el marco

No conocemos ejemplos bien descritos de un discurso de políticas que refleje el marco. Pero queremos imaginar aquí cómo sería y cuáles serían las condiciones que lo favorecieran. Al parecer hay dos visiones principales de los participantes en una controversia de políticas. Podemos imaginarlos como *investigadores* que consideran la situación de una política cual objeto de pensamiento, experimentación o análisis. Desde esta posición aventajada, sus intereses propios son fuentes potenciales de distorsión. O podemos imaginar a los participantes como *partes interesadas* —patrocinadores o partidarios de marcos— que utilizan la investigación para favorecer sus propios intereses. En el primer caso, vemos la política como una copa sobrepuesta a la investigación. En el segundo caso, la investigación surge de un proceso básicamente político⁹.

De acuerdo con la segunda visión, los marcos dan una forma cognoscitiva a los intereses que surgen de roles sociales, posiciones, membresías e historias. La situación paradigmática es la de dos partes empeñadas en una lucha por el dominio de algo, lucha que necesariamente toma la forma de un juego de ganar y perder. Lo que está en juego en la lucha es ¿quién sostiene que debe tener la prioridad? Según esto, la función de la reflexión consiste en resolver la disputa, o tal vez en transformar la disputa en tal forma que sea más reductible a una solución satisfactoria. La actitud subyacente que apela a la reflexión es algo como «¡seamos racionales!».

Según la primera opinión, los participantes parecen miembros de un sistema social cooperativo. Se enfrentan a una situación problemática común, tienen un interés compartido en reenmarcar y resolver, aunque acaso puedan ver esto, al principio, de maneras distintas. La situación paradigmática es la de dos personas que persiguen una tarea común y se enfrentan a un obstáculo puesto a su empeño compartido. La función de la reflexión consiste en facilitar su investigación del problema común; la actitud subyacente es algo como «pensemos en lo que estamos haciendo».

Llamaremos a estas las visiones *política y cooperativa* de la controversia de políticas. Según ambas opiniones existen conflicto e investigación, pero se los concibe de diferentes maneras. Si se ve la controversia de las políticas como básicamente cooperativa, el llamado a la reflexión es un llamado a hacer que la investigación compartida sea más significativa y eficaz. Si se ve la controversia de las políticas como política, el llamado a la reflexión es un llamado al interés común de los participantes en minimizar o al menos en compartir los costos experimentados en el juego de ganar o perder: en el mejor de los casos, transformar el conflicto en cooperación. En el peor, permitir al menos a los participantes llegar a una solución satisfactoria de su disputa.

Estos rasgos generales del discurso del marco pueden relacionarse ahora con nuestro anterior análisis de la búsqueda hecha por Rorty de un discurso normal en relación con el cual podamos estudiar o influir sobre el discurso anormal. La traducción a través de marcos en conflicto es un modo de pensar acerca de la relación del discurso anormal con el discurso normal. Por ejemplo, podemos enfocar la labor de quienes se han arrogado como misión especial la tarea de tender un puente entre el discurso normal y el anormal no solo en la política social, sino también en la ciencia. Este punto de vista, basado en la misión especial del traductor, parece perder de vista el punto esencial de que las controversias de políticas son sustantivas: acerca de premisas y axiomas, como lo ha señalado Simon (1983) y no solo acerca de las reglas para llegar a un acuerdo. El significado precede a los hechos y solo tiene sentido en relación con los propósitos situados en diferentes marcos. De hecho, el conflicto de los marcos ha contribuido a la reciente reformulación del papel de la ciencia en las controversias de políticas. Hay crecientes dudas acerca de la neutralidad de la ciencia como agente en los debates causados por el enmarque en las cuestiones de políticas. Si ya no se puede confiar en que la ciencia mediará en los conflictos sobre marcos, habrá el riesgo de aumentar el nivel del discurso anormal.

Por otra parte, las personas parecen vivir en un mundo que es a la vez consensual y conflictivo. Coexisten el discurso normal y el anormal. En algunos aspectos, la gente parece compartir los metamarcos: en la ciencia, la fe en la razón y la evidencia; en las políticas, la fe en el capitalismo democrático. Cuanto más abstracto sea el principio, más alto será el nivel

del acuerdo. Surge la discordia cuando nos volvemos explícitos, es decir, cuando significado y propósito cobran vida en situaciones reales. Por ejemplo, aunque todos los disputantes estén de acuerdo en la democracia, pueden estar en desacuerdo acerca del nivel de gobierno que debe gobernar, el tipo de democracia que se debe subrayar o cómo debe actuar el gobierno para que las instituciones se enfrenten a las fluctuaciones económicas y para proteger a los individuos contra la incertidumbre. Es en las situaciones concretas (donde se unen la política y el mercado) donde se hace más evidente el conflicto de marcos. Sin embargo, aun entonces hay un hogar simbólico al que los disputantes pueden retirarse para buscar consenso en un nivel superior de abstracción, donde la vaguedad resulta unificadora y donde hay un compromiso común con cierto tipo de discurso normal. En realidad, el hecho mismo de participar con otros en el discurso presupone cierto marco de referencia compartido, aun si solo es un conocimiento compartido de, o una lealtad con, reglas del juego socialmente legitimadas. Por su controversia misma, los actores se declaran miembros de un sistema social en que se comprenden y obedecen unas reglas tácitas de esta clase de lucha.

Vemos así que hay un metanivel donde hay un tipo de consenso que es afín al orden social. Pero esta casiconsensualidad puede dificultar más aún la tarea del discurso. Al nivel superior, los disputantes pueden hablar sin oírse unos a otros, inconscientes de su verdadero desacuerdo. Solo en la diaria tarea de hacer y analizar las políticas se manifiesta el choque entre los diversos marcos; en otros niveles puede ser desesperantemente elusivo.

Un ejemplo de ello fue el debate acerca de la política sobre las armas nucleares. El discurso normal enmarca el debate en términos de un limitado conjunto de opciones. Se trata de las opciones pensables y discutibles que dependen de un conjunto compartido de axiomas y premisas, el más importante de los cuales es la potencial amenaza de la Rusia soviética a la seguridad norteamericana. Según esta opinión, es necesaria cierta estrategia defensiva, y el desarme unilateral no es una opción en que siquiera pueda pensarse. Otro marco considera la cuestión en términos de aspiración humana y la aniquilación que seguiría a una guerra nuclear. En ese marco, las consideraciones estratégicas y militares están proscritas.

No se hace un serio intento de comunicación entre los dos marcos. La estrategia de la protesta se pone en términos de drama personal, expresado apelando al futuro bienestar de los niños del mundo. Pero, desde luego, no se excluye por completo cierto elemento de comunicación. Entre estos marcos existe todavía un discurso vacilante. Queda cierta esperanza de convertir el discurso anormal en normal, tal vez convirtiendo la controversia entre los marcos en un desacuerdo dentro de un marco. Pero los llamados a la ciencia, como en los recientes debates por el «invierno nuclear», fácilmente son presa de una reconversión del discurso normal en anormal; ambos bandos ofrecen interpretaciones predeciblemente conflictivas de «los mismos datos».

Los conflictos sobre políticas a veces se resuelven por medio de procesos interactivos que no entrañan un conflicto de marcos. Los participantes pueden entablar una lucha en cualquiera de los contextos institucionales de lucha que les ofrezca su sociedad, y uno u otro bando de la disputa podrá ganar. O bien, si no surge un ganador indiscutible, las partes de la disputa podrán negociar un acuerdo en que convendrán en cualquiera de los contextos institucionales de negociación que la sociedad les ofrezca.

Cuando el discurso es en el modo político, y cuando los objetos de reflexión son las posiciones de los participantes o los conflictos que hay entre ellos, entonces la labor de la reflexión puede consistir en la adjudicación, negociación, ocultación o administración del conflicto. Pero las controversias más reacias de políticas, como la controversia por las armas nucleares, no son reductibles a este tipo de acuerdo. No hay «vencedores». O bien las victorias temporales dejan sin resolver la controversia básica. O los acuerdos negociados en un contexto local particular o en torno a una cuestión particular, como en el caso de la reciente historia de la política de bienestar, no logran un consenso político estable. Nosotros afirmamos que cualquier proceso tendiente a la resolución de enconadas controversias de políticas deberá incluir sus conflictos entre marcos subyacentes, y que los analistas de las políticas pueden desempeñar un papel importante en este proceso.

Los analistas de las políticas que criticaran los marcos descubrirían los múltiples marcos conflictivos que participan en cualquier disputa de políticas. Averiguarían las fuentes de los marcos conflictivos en las

historias, roles, contextos institucionales e intereses de los participantes. Explorarían las ambigüedades e incongruencias contenidas en ellos y las consecuencias que podría tener su uso.

Pero ese análisis, cuando se emprende desde la perspectiva remota de un especialista universitario, tendría un potencial limitado para contribuir a elaborar una resolución mutuamente satisfactoria de las controversias de políticas. En el mejor de los casos, daría a los participantes una mayor percatación de sus modos de enmarcar cada problema. No conduciría *per se* al proceso que Habermas (1968) ha descrito como «autorreflexión crítica que contribuye al consenso político».

Consideramos necesario explorar el espacio que yace entre las luchas políticas y las negociaciones características del ordinario debate político y el análisis —crítico de marcos— de un estudioso protegido y aislado en su universidad. Desde esta perspectiva, veríamos a los participantes en una controversia de políticas como agentes-experimentales (véase Vickers, 1975), en transacción (véase Dewey y Bentley, 1949) con las situaciones de las que forman parte. Acaso logren modificar sus situaciones, que también pueden cambiar como resultado de las modificaciones operadas en algún más vasto contexto social, y pueden ser modificadas por los cambios de su situación, algunos de los cuales tal vez ellos mismos ayudaran a producir. Estas condiciones transaccionales son básicas para el desenvolvimiento de la historia de una situación de políticas. En términos esquemáticos:

Las personas se encuentran en presencia de esta situación, de la cual cobran conocimiento.

A partir del enmarque inicial de la situación llevado a cabo por las personas, estas emprenden una acción

Que contribuye al cambio de la situación,

Por la cual son afectadas,

Y las personas cobran conocimiento de este cambio,

Sobre la base del cual vuelven a actuar

Y se actúa sobre ellas.

En el curso de la *conversación* de los participantes, con su situación pueden ocurrir el reflejo del marco y el resultante cambio de marco. Es importante

observar que no es probable que ocurra semejante cambio «en nuestra mente» antes de emprender una acción. Es más probable que ocurra conforme se vaya captando y se responda a ella.

Un cambio de marco puede no tener ningún sentido. Sencillamente podemos *encontrarnos* teniendo que pensar en las cosas de otra manera. Como ya lo subrayó Festinger (1957), nuestro horror a la «disonancia cognoscitiva» —la discordia de nuestras creencias con nuestra conducta real— puede hacer que, gradual y tácitamente, ajustemos nuestras creencias para dar acomodo a los cambios de nuestras situaciones. Esto significa que continúa una labor cognoscitiva de cierta índole, pero sin una crítica o un control consciente. Sin embargo, habiendo tenido que hacer ese cambio, bien podemos esforzarnos conscientemente por justificarlo o, si descubrimos que no se lo puede justificar, podemos esforzarnos conscientemente por anularlo.

Un cambio de marco también puede ser consciente. En un momento u otro de la conversación de los participantes con una situación de política —típicamente, en alguna «coyuntura» de la dialéctica de las políticas—, pueden, en el discurso de uno con otro, someter sus marcos en conflicto a un pensamiento y un control conscientes y, como consecuencia parcial, reenmarcar la situación.

Cuando el discurso refleja el marco, la labor de la reflexión consiste ante todo en sacar a la superficie marcos conflictivos, identificar sus fuentes y consecuencias, y luego traducir, reestructurar, integrar, convertir o elegir.

La labor cognoscitiva del discurso que refleja el marco depende de diversas condiciones contextuales. Depende de la *actitud* de los participantes: su distancia relativa con respecto a sus materiales y con respecto a los procesos en que participan, sus actitudes hacia la incertidumbre que sigue al *desenmarque*, su disposición a correr riesgos cognoscitivos, su inclinación a aceptar las opiniones de otros, y su capacidad para la doble visión que puede dar a las personas el don de combinar la defensa de un marco con la investigación de otros. Por todas estas razones, la labor del reflejo de marcos es tanto afectiva como cognoscitiva; incluye sentimientos y análisis de sentimientos.

Por último, el reflejo de los marcos probablemente sea inseparable del reflejo del contexto. Como lo ha señalado Kuhn (1977), acaso los científicos logren la tarea cognoscitiva de traducir de un paradigma a otro, pero, ¿cómo haremos para que se sienten juntos a discutirlo? El reflejo recíproco de los marcos depende de la creación de un mundo conductual que conduzca a él.

Todos estos factores merecen más investigación de las condiciones para el discurso de políticas que refleje los marcos. Una investigación crítica del conflicto de marcos y del reflejo de marcos en las arenas interrelacionadas del cambio institucional, la investigación erudita o científica y la vida personal, plantea un fundamental desafío analítico a la investigación de políticas contemporáneas. Sin embargo, por ese mismo hecho, contiene el potencial necesario para extender considerablemente la investigación más allá de sus actuales límites y, así, para iluminar las carreras de las más enconadas controversias de políticas.

Notas

* Título original "Social Sciences and Modern States". Primera edición en inglés ©1991 por University of Cambridge Press. Traducción autorizada al español "El discurso que refleja su marco" en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.), "Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas", pp. 327-356, Fondo de Cultura Económica, México, 1999. Reproducción autorizada.

¹ Véase Majone, en el capítulo XIII. En su ensayo, Majone se explaya sobre otra ruta de la investigación al mismo problema, a saber: una investigación de hasta qué punto la epistemología filosófica puede servir útilmente como analogía y vehículo para el análisis de problemas de políticas, incluyendo las más profundamente arraigadas controversias de políticas.

² Una vez más, encontramos una notable complementariedad entre nuestro enfoque y el de otra colaboración a este volumen, a saber, la de Wittrock *et al.*, capítulo II. Mientras que el discurso de políticas se refiere a un tipo de proceso de conversión, a saber: aquel por el cual las interacciones directas entre diversos intereses se convierten en problemas de política, este enfoca —para emplear su término— la estructuración del discurso. En tales procesos, problemas de política de particular importancia y persistencia pueden convertirse —y fusionarse— en discursos intelectuales institucionalmente reproducidos, que a su vez pueden servir como depósitos para su uso intermitente en varias controversias, ya sea en una función de ilustración o de defensa (véase también Weiss [capítulo XIV] y Wittrock

[capítulo XV], quienes elaboran una serie de modelos precisamente de estos tipos de utilización del conocimiento).

3. Esta referencia fue sugerida por Lisa Peattie.
4. Véase también Majone (capítulo XIII), Weiss (capítulo XIV) y Wittrock (capítulo XV), quienes de diferentes maneras enfocan este dilema.
5. En el capítulo V de este volumen, Smith demuestra una pauta similar de continuidad a lo largo del espectro partidista con respecto a la política británica de la ciencia social en los muy diferentes contextos de las políticas de los sesenta y los setenta.
6. Schwartzman, en el capítulo XI, ofrece un relato fascinante de las maneras en que ocurren los cambios en los discursos de políticas sobre largos periodos, inmediatamente detrás de los giros en las posturas de aquellos que portan las pretensiones y transiciones del conocimiento en sus dominios institucionales y profesionales.
7. La descripción del discurso de políticas que aquí damos es una modificación de una antigua descripción, hecha por Schon, de los procesos por los cuales las ideas de políticas cobran aceptación (Schon, 1971).
8. Gran parte de este análisis se basa en conversaciones informales sostenidas entre Martin Rein y Hugh Hecló.
9. Desde luego, este mismo problema es examinado, con cierta extensión, por varios de los colaboradores, tal vez más notablemente por Schwartzman (capítulo XI) y por Weiss (capítulo XIV).

Referencias

- Argyris, Chris, y Donald A. Schon, 1974. *Theory in Practice*. San Francisco, Jossey-Bass.
- Bateson, Gregory, 1972. *Steps to an Ecology of Mind*. Nueva York, Ballantine Books.
- Dewey, John, y Arthur Bentley, 1949. *Knowing and the Known*. Boston, Beacon Press.
- Festinger, Leon, 1957. *A Theory of Cognitive Dissonance*. Evanston, IL, Row Peterson and Co.
- Friedan, Betty, 1964. *The Feminine Mystique*. Nueva York, Norton Press.
- Gamson, William A., y Kathryn Lasch. De próxima aparición. «The Political Culture of Affirmative Action». Mimeografiado.
- Geertz, Clifford, 1983. *Local Knowledge*. Nueva York, Basic Books, Inc.
- Goffman, Erving, 1974. *Frame Analysis: An Essay on the Organization of Experience*, Cambridge, MA, Harvard University Press.
- Goodman, Nelson, 1978. *Ways of Worldmaking*. Indianapolis, Hackett.
- Habermas, Jürgen, 1968. *Knowledge and Human Interests*. Boston, Beacon Press. [Traducción al español en Taurus].

- Hecló, Hugh, 1985. «Issue Networks and the Executive Establishment» (mimeografiado).
- Innis, Harold, 1951. *Bias of Communication*. University of Toronto Press.
- Kahneman, Dan, y Amos Tversky, 1974. «Judgment under Uncertainty: Heuristics and Biases». *Science* 185, pp. 1124-1131.
- Kuhn, Thomas, 1962. *The Structure of Scientific Revolution*. University of Chicago Press. [Hay traducción al español del FCE].
- Mannheim, Karl, 1936. *Ideology and Utopia*. Nueva York, Harcourt, Brace and Co. [Hay traducción al español del FCE].
- March, James, y Johan P. Olsen, 1975. «The Uncertainty of the Past, Organizational Learning Under Ambiguity». *European Journal of Political Research* 3 (marzo), pp. 147-171.
- Mead, Lawrence, 1985. *Beyond Entitlement*, Nueva York, Free Press.
- Minsky, Marvin, 1978. «Frames». *Artificial Intelligence Laboratory Memorandum*, MIT, Cambridge, MA.
- OECD (Organization for Economic Co-operation and Development), 1981. *The Welfare State Crisis*. París, OECD.
- , 1985. *Social Expenditures 1960-1980: Problems of Growth and Control*. París, OECD.
- Parsons, Talcott, 1967. *Social Systems and the Evolution of Action Theory*. Nueva York, Free Press.
- Piven, Francis, y Richard Cloward, 1971. *Regulating the Poor: The Functions of Public Welfare*. Nueva York, Pantheon Books.
- Rein, Martin, y Donald A. Schon, 1977. «Problem Setting in Policy Research». En *Using Social Research in Public Policy Making*, pp. 235-251. Ed. Carol H. Weiss, Lexington, MA, Lexington Books.
- Rorty, Richard, 1979. *Philosophy and the Mirror of Nature*. Nueva York, Princeton University Press. [Hay traducción al español].
- Schon, Donald A., 1971. *Beyond the Stable State*. Nueva York, Random House.
- Schutz, Alfred, y Thomas Luckmann, 1981. *The Structures of the Life World*. Evanston, IL, Northwestern University Press (publicado por primera vez en 1973). [Traducido al español por Amorrorrtu].
- Simon, Herbert, 1988. *Reason in Human Affairs*. Londres, Basil Blackwell.
- Stevens, Wallace, 1954. *The Collected Poems of Wallace Stevens*. Nueva York, Alfred A. Knopf.
- Vickers, Geoffrey, 1975. «Social and Institutional Reality». Cambridge, MA (mimeografiado).
- Wollmann, Hellmut, 1984. «Policy Analysis: Some Observations on the West German Scene». *Policy Sciences*, pp. 27-47.

Programas de investigación y programas de acción o ¿puede la investigación de políticas aprender de la filosofía de la ciencia?*

Giandomenico Majone

En el capítulo anterior, Rein y Schon se basaron en teorías del discurso y en el constructivismo para ilustrar cuestiones de políticas. Yo deseo volverme hacia un lugar al parecer más inesperado para buscar en él ideas acerca de la política pública: me refiero a la filosofía de la ciencia. Si es verdad que el conocimiento científico es un conocimiento de sentido común «con mayúsculas» (como constantemente lo subraya Karl Popper), será al menos posible que los estudiosos de la política pública descubran algo digno de aprender en los recientes debates metodológicos sostenidos entre historiadores y filósofos de la ciencia. Sea como fuere, esta es la intrigante posibilidad que deseo explorar en este capítulo. Más precisamente, sostendré que los modelos conceptuales creados por Popper y por Imre Lakatos para explicar y evaluar el desarrollo del conocimiento científico son aplicables (al menos heurísticamente) al estudio del desarrollo de las políticas y a la práctica de la evaluación de políticas.

En años recientes, un número cada vez mayor de estudiosos se ha convencido de que no es posible comprender las políticas públicas solo con base en conceptos como poder, influencia, interés, presión y decisión. Las ideas y los argumentos, lejos de ser simples racionalizaciones o embellecimientos retóricos, son partes integrales de la hechura de las políticas (véase también Weiss, capítulo XIV), y el desarrollo de las políticas siempre va acompañado de un proceso paralelo de innovación conceptual, debate y persuasión. Por ello, todo relato adecuado de la evolución de las políticas

públicas debe tener en cuenta el conocimiento y las ideas, así como la política y la economía. Lo difícil es lograr que ambos conjuntos de factores —el orden de las ideas y el orden de los hechos— se unan para ofrecer un solo cuadro coherente. En este capítulo nos basamos en los modos en que los órdenes de las ideas han sido concebidos por los filósofos de la ciencia, y tratamos de mostrar cómo, por vía del razonamiento analógico, pueden detectarse órdenes similares en los hechos que ocurren, por medio de una actividad humana consciente y racionalmente orientada en la sociedad y en el Estado moderno: los hechos a que nos hemos acostumbrado a llamar *elaboración de las políticas*.

Programas de investigación científica

La metodología de Lakatos para enfocar los programas de investigación científica es al mismo tiempo una «lógica de la evaluación» —una metodología normativa para evaluar el progreso científico— y una teoría histórica del desarrollo científico. El problema histórico que Lakatos intenta resolver es la coexistencia del cambio y la continuidad en la ciencia. Filosofías anteriores parecían implicar que los científicos (al menos los científicos con probidad profesional) descartan sus teorías al primer contraejemplo. Pero la historia muestra que hasta los mejores científicos suelen pasar por alto los contraejemplos; continúan elaborando y practicando sus teorías a pesar de obtener resultados experimentales desfavorables. En realidad, esta actitud de aferrarse el mayor tiempo posible a una teoría tiene sentido, ya que, como lo observa Kuhn (1970: 145), «si todos y cada uno de los fracasos fueran motivo para el rechazo de una teoría, habría que rechazar todas las teorías en todo momento». Un hombre o mujer de ciencia que abandone su hipótesis a los primeros resultados negativos no logrará descubrir su solidez.

La tenacidad (o el dogmatismo) en aferrarse a una teoría ante una evidencia negativa, que Kuhn considera como rasgo fundamental de una ciencia «normal», puede explicar la continuidad. Pero entonces, ¿cómo explicamos el desarrollo científico? Kuhn emplea la metáfora de la revolución, mientras que la metodología de Lakatos nos ofrece una explicación evolutiva de la continuidad y a la vez del cambio. Primero, Lakatos indica que, como las teorías no existen aisladas, sino que están

incluidas en un sistema de ideas más o menos integrado, la unidad apropiada del análisis no es una teoría aislada, sino toda una estrategia de investigación o, como él dice, un programa de investigación científica. Luego, los programas de investigación se descomponen en una parte rígida y una flexible. La parte rígida incluye un «núcleo duro»: las suposiciones básicas aceptadas y vueltas provisionalmente irrefutables por una decisión metodológica de los científicos que apoyen el programa. Una «heurística positiva», que consiste en un conjunto de sugerencias sobre cómo elaborar hipótesis auxiliares y variantes refutables del programa de investigación; y una «heurística negativa», que indicará los caminos de investigación que deban evitarse. Por ejemplo, el programa cartesiano de investigación —el universo es una enorme maquinaria de relojería, en que el empuje es la única causa del movimiento— nos dice que busquemos más allá de todos los fenómenos naturales, incluso la vida, para encontrar explicaciones mecanicistas (heurística positiva), mientras excluimos la acción newtoniana a cierta distancia (heurística negativa).

La economía neoclásica y la marxista son ejemplos de programas de investigación que compiten en las ciencias sociales. El núcleo de la economía neoclásica es el concepto de que la economía se compone de agentes libres que desempeñan diferentes funciones, pero comparten el objetivo de maximizar la utilidad individual. La idea de que la sociedad está dividida en clases sociales en mutua competencia se rechaza por considerarse que no es correcta ni analíticamente útil. Del lado positivo, la economía se concibe como el estudio de la asignación racional de recursos escasos entre usos en competencia. Este estudio (se sostiene) puede efectuarse a un nivel abstracto cualesquiera que sean las condiciones históricas específicas, ya que la meta de maximizar la utilidad individual tiene validez universal.

La idea central de la economía marxista, en cambio, es que la sociedad está dividida en clases que buscan metas diferentes y conflictivas. Según la heurística negativa del programa de investigación marxista, no se pueden explicar los fenómenos económicos en función de conductas individuales, y es imposible definir una norma de eficiencia económica que sea válida para toda una sociedad. La heurística positiva del programa llama la atención hacia la lucha de clases, la explotación y una teoría del valor trabajo como categorías básicas del análisis. Dado que la formación

de clases y la lucha de clases difieren en naturaleza y caracterizan a tipos particulares de sociedad, por ejemplo, el capitalismo, una teoría económica del capitalismo solo puede ser válida para esta fase particular de la historia de la humanidad.

La parte flexible de un programa de investigación científica es su «cinturón de seguridad», es decir, el conjunto de hipótesis auxiliares y predicciones específicas que se puede someter a la prueba de la evidencia empírica sin comprometer la integridad del núcleo central. El cinturón protector está en un continuo fluir conforme sus elementos se adaptan, refinan y descartan; en este proceso, el programa de investigación se gana su reputación científica. Pero un programa de investigación no es científico de una vez por todas. Deja de serlo cuando se vuelve incapaz de producir nuevas ideas, de predecir acontecimientos inesperados y de dar acomodo a nuevos hechos, salvo sobre una base *ad hoc*. Así, según Lakatos, la manera más fructífera de concebir la historia de la ciencia no es como un avance firme, interrumpido de cuando en cuando por una revolución científica, a la manera de Kuhn, sino como una sucesión de programas de investigación que están constantemente descartándose unos a otros con teorías de un contenido empírico cada vez mayor (Lakatos, 1971; Blaug, 1975).

Las políticas como «programas de acción»

Veamos ahora si la metodología de Lakatos puede hacer algunas sugerencias útiles para el estudio de la política pública. En cualquier ámbito de la actividad gubernamental (defensa, energía, educación, etc.), ciertos conjuntos de programas caen dentro de pautas que muestran una continuidad reconocible mientras que, al mismo tiempo, sufren constantes cambios. A estas *gestalts* relativamente estables, cuyas propiedades no son derivables de sus elementos programáticos constitutivos por simple adición, las llamo *programas de acción* o *de políticas*, por analogía con los programas de investigación científica de Lakatos (el término *programa de acción* fue introducido por Boothroyd, 1978, por referencia explícita a la terminología de Lakatos).

Para el analista, identificar los principios integrantes de un programa de acción o de políticas es un problema tan grande como lo es comprender

cómo cambian ciertos elementos programáticos en particular. Como lo ha observado Heclo (1972: 85), «las políticas no parecen ser un fenómeno que se defina a sí mismo; son una categoría analítica cuyos contenidos son identificados por el analista, no por los dichos del responsable de las políticas, ni por las piezas de legislación o de la administración». Pero si se quiere comprender las políticas en su doble aspecto de cambio y continuidad, hay que refinar más esta categoría analítica.

Si una política no va a ser abandonada a las primeras señales de dificultad o sacrificada para obtener ventajas a corto plazo, deberá basarse en algunos principios y compromisos vinculativos y generalmente aceptados. Como en el caso de la formación de instituciones, la creación de normas o estándares que solo se puedan cambiar (si acaso) al riesgo de graves crisis internas, es un paso crucial en la elaboración de las políticas.

Normas, estándares, ideologías y creencias son parte de lo que podemos llamar el *núcleo de las políticas*. Sin embargo, tales principios generales no bastan para definir una política; también se necesitan estrategias que traduzcan los principios en práctica. Por tanto, el núcleo de las políticas debe incluir criterios para distinguir entre los «problemas» y «el modo como son las cosas», entre los cursos de acción aceptables y los inaceptables. En la terminología de Lakatos, tales normas son la *heurística positiva* y la *heurística negativa* de las políticas. Representan el «hay que hacer» y el «no hay que hacer» del programa de acción.

Por último, los planes, programas y acciones con que se pretende poner en vigor los principios generales del núcleo son su *cinturón protector*. En realidad, un núcleo de políticas tendrá, característicamente, varios cinturones de seguridad a los que podemos imaginar como una secuencia de círculos concéntricos, con el núcleo como su centro común. La razón de imaginarnos así los programas es que cuanto más nos acercamos al núcleo desplazándonos a lo largo de la secuencia de cinturones protectores, mayor será la resistencia a los cambios (en la siguiente sección ofrecemos un ejemplo).

La distinción entre el núcleo y el cinturón protector subraya una función importante, pero a menudo pasada por alto, de los programas operacionales

y las disposiciones administrativas. Para desempeñar su papel normativo, el núcleo debe estar protegido en lo que sea posible contra toda crítica y contra cambios demasiado frecuentes o profundos. La política exterior *bipartidista* es ejemplo de un acuerdo común, si bien a menudo tácito, entre participantes clave en el proceso de las políticas, con la intención de hacer «provisionalmente irrefutables» ciertos principios nucleares. Esto significa que el debate, la controversia y las correcciones deben ir redirigidos hacia los programas operacionales y las disposiciones institucionales que forman el cinturón de seguridad del núcleo. La necesidad de adaptar los mecanismos administrativos particulares según los cuales operan las políticas, a unas condiciones económicas, sociales y políticas en eterno cambio mantienen a ese cinturón en constante fluir, pero tales modificaciones no suelen afectar el núcleo de las políticas. Desde luego, una vez que la base del consenso ha socavado el núcleo, con el tiempo caerá, pero esto significará un cambio importante de las políticas: en cierto sentido, una revolución.

¿Cómo identificamos los diversos elementos de un programa de acción? Como en la interpretación histórica, no hay reglas ya hechas en toda forma. Cualquier reconstrucción racional de los asuntos humanos está expuesta a cierto número de potenciales trampas: la mirada retrospectiva, los sesgos, la excesiva simplificación. Pero esto no es decir que un hábil analista pueda transformar cualquier colección de piezas de legislación o de administración en una política, así como el historiador no es libre de descubrir en los hechos pasados el desenvolvimiento de su teoría predilecta. Hay, primero, evidencia documental (leyes, debates legislativos, decisiones judiciales, historias y registros oficiales, etc.) contra los cuales se puede compulsar la reconstrucción propuesta. En segundo lugar, la importancia operacional de la heurística positiva y negativa debe hacerse evidente en las doctrinas programáticas, los procedimientos operativos habituales, los instrumentos de política preferidos y las alternativas rechazadas o «no decisiones». Otra fuente relevante para una comprensión clara de las cosas estriba en las opiniones y acciones de los partidarios y a la vez de los adversarios de una política determinada. Por lo general, los actores de las políticas no apoyan ni se oponen con igual vigor a todos los cambios propuestos en las políticas existentes; sus reacciones diferenciadas pueden ofrecernos claves importantes acerca de la relativa significación de los diversos elementos del programa de acción.

También importa reconocer cuándo un conjunto de actividades no llega a ser una política. A manera de ilustración, mencionaremos dos casos extremos. Primero, los programas y las decisiones que el analista observa pueden ser tan solo los restos dispersos de lo que antes fue una política en toda forma: fragmentos de un cinturón de seguridad que se ha convertido en cáscara vacía. La metodología de Lakatos incluye el concepto de programas de investigación «degenerativos»: programas que son cada vez menos capaces de predecir hechos inesperados y que, si acaso, solo pueden dar acomodo sobre una base *ad hoc* a cualesquiera hechos nuevos que ocurran. Estos programas de investigación degenerativos se vuelven demasiado engorrosos para conservar su atractivo intelectual y conquistar talentos científicos nuevos. De manera similar, una política que ya no logra producir nuevas ideas interesantes, que continúa combatiendo en las batallas de ayer en lugar de atacar los problemas de hoy, con el tiempo perderá credibilidad intelectual y apoyo político. Su núcleo se atrofiará hasta ser remplazado, pero partes del cinturón de seguridad podrán seguir existiendo durante algún tiempo. Recientes acontecimientos económicos ocurridos en muchos países occidentales ejemplifican el proceso de progresiva decadencia de un núcleo de políticas: el keynesianismo ortodoxo. Queda por ver si alguna versión del monetarismo o de la economía del lado de la oferta logrará remplazar al antiguo núcleo.

El segundo caso en que puede afirmarse con toda confianza que no existe una política reconocible es bastante directo: o bien un núcleo no ha surgido aún (este parece ser el caso de la «política» energética de muchos países actuales), o bien le faltan algunos de sus elementos esenciales. Por ejemplo, a comienzos de los cuarenta, Lippmann (1943: 3) sostuvo que durante casi cincuenta años, los Estados Unidos no habían poseído una «política exterior decidida y generalmente aceptada». Su argumento, según la terminología de este capítulo, es que el núcleo de la política exterior norteamericana estuvo incompleto en el periodo transcurrido entre el principio del siglo y el comienzo de los cuarenta. Sí incluyó algunos principios generales y compromisos de largo alcance (dentro de los límites continentales de los Estados Unidos, en el hemisferio occidental, en el Pacífico), pero no instrumentos y estrategias creíbles (heurística) para poner en vigencia esos principios y compromisos.

Anatomía de dos programas de acción

Habiendo introducido toda la nueva terminología que necesitamos por el momento, podemos ahora examinar algunos ejemplos. Empecemos por el desarrollo de la política de seguridad social en los Estados Unidos. Uno de los rasgos más sorprendentes de este desarrollo es que, aun cuando el sistema ha pasado por constantes cambios incrementales desde su creación en 1935, los cambios han seguido rutas bien definidas por doctrinas programáticas. Desde el principio mismo, la administración superior explicó con toda claridad sus primeros principios: el programa debía ser por contribución «universal» y obligatorio; debía ser de alcance nacional y administrado por el gobierno federal. Las prestaciones debían estar relacionadas con los salarios; habiendo pagado sus aportaciones, los asegurados o sus dependientes debían recibir sus beneficios como una cuestión de derecho. Así, pues, estos pueden considerarse como los principios nucleares de la política de seguridad social (véase también Bulmer, capítulo VI).

La principal heurística positiva subraya el método de seguro social como técnica preferida de elaboración de las políticas. Según este principio, no solo es obligatoria la participación, sino que sus beneficios quedan prescritos por ley y no por contrato, y el sistema debe ser moderadamente redistributivo (en favor de los grupos de escasos ingresos, de los asegurados con muchos dependientes y de los participantes que ya eran viejos cuando se inició el sistema). Esta heurística positiva ha sido muy eficiente. Según Derthick (1979), la técnica prevaleciente del análisis de políticas en la Administración de la Seguridad Social consistió en identificar un problema social, como la falta de atención a la salud, y en crear argumentos y métodos para hacerle frente por medio de la seguridad social.

La principal heurística negativa rechaza —lo cual no es de sorprender— el principio del seguro privado (según el cual la protección es voluntaria y la ofrecen muchos aseguradores en competencia mutua, mientras el costo de cada protección queda determinado sobre una base estrictamente actuarial), pero también se opone al asistencialismo público, los estudios socioeconómicos para determinar el grado de necesidad y otros enfoques «benefactores» a los problemas sociales.

El cinturón protector del programa de acción de seguridad social contiene muchos elementos. Sin embargo, los ejecutivos de programas y los principales partidarios políticos no consideran que todos los elementos programáticos tengan igual importancia, es decir, que estén igualmente cerca del núcleo. Por ejemplo, el Seguro de Vejez y Sobrevivientes (OASI) ocupa un lugar mucho más central y, por tanto, está mucho más cuidadosamente protegido que, por ejemplo, la Asistencia a la Vejez (OAA). Según esta métrica (que mide la distancia del núcleo), el Seguro de Incapacidad y Medicare ocupan una posición intermedia entre OASI y OAA, estando mucho más cerca del primer programa que del último.

De este modo, la política de seguridad social en los Estados Unidos (como en muchos países europeos) tiene un núcleo reconocible, el cual contiene un conjunto claramente definido de principios y reglas metodológicas. Sin embargo, al mismo tiempo, este núcleo permite operar una gran variedad de opciones (Derthick, 1979: 21-23). El programa no tiene que ser financiado solo mediante contribuciones, ni las personas necesitan aportar mucho y durante largo tiempo para tener derecho a sus beneficios; estos deben estar relacionados con los salarios, pero la naturaleza exacta de la relación puede determinarse de muy distintas maneras. Hasta el uso de ingresos generales sería compatible con los principios nucleares, y en realidad este enfoque muy pronto fue propuesto por los ejecutivos del programa. En suma, el cinturón protector de la política ha demostrado ser notablemente flexible, y esta flexibilidad ha contribuido grandemente al éxito político del programa, al menos hasta hace poco.

Por cierto que el ejemplo de la seguridad social sugiere que la flexibilidad táctica (en el cinturón protector), lejos de ser incompatible con un núcleo de políticas claramente definido, en realidad puede ser ayudada por él. Por ejemplo, a menudo se ha observado que los partidos políticos con marcados compromisos ideológicos suelen ser más propensos a las adaptaciones pragmáticas y los acuerdos que los partidos que no tienen normas claras con las cuales distinguir lo que es esencial de lo que es negociable.

La política de salud británica, de acuerdo con el Servicio Nacional de Salud, nos ofrece otro ejemplo instructivo. Creado en 1946, el Servicio ha pasado por buen número de cambios administrativos y de organización,

pero ha mantenido su compromiso básico con el principio nuclear de que se deben distribuir los servicios médicos de acuerdo con la necesidad más que con la capacidad para pagar. La heurística positiva de las políticas prescribe estrategias que son compatibles con el principio básico, como el financiamiento por medio de impuestos generales y la nacionalización de los hospitales privados, mientras que la heurística negativa excluye otros enfoques, como los cargos a los usuarios (con excepción de algunos servicios de salud no esenciales). Incluso las propuestas de los «bonos de salud» han sido rechazadas hasta ahora por los gobiernos conservadores y por los laboristas, por considerarlas incompatibles con el principio de elección colectiva en cuestiones de salud.

Los problemas de demanda excesiva creados por la abolición de las barreras financieras no han sido atacados reintroduciendo mecanismos de precios (cuota por servicio), sino fijando límites presupuestarios a los gastos totales en salud, mediante controles administrativos y mejora en la eficiencia de los hospitales y otros ajustes del cinturón de seguridad. De hecho, el debate acerca del núcleo del Servicio Nacional de Salud, en contraste con su cinturón, por lo general se ha considerado impertinente. Esto queda en claro en los siguientes comentarios de Blaug (1970: 324):

Nos guste o no, el Servicio Nacional de Salud Británico reemplazó efectivamente la opción individual en la distribución de los servicios de salud por la elección colectiva. Por ello, son totalmente impertinentes en Inglaterra los argumentos acerca de la «falla del mercado» para justificar la propiedad o el financiamiento gubernamental, a menos, desde luego, que la tesis sea que deben volverse pertinentes devolviendo la salubridad al mecanismo del mercado. Diríase que hoy existe un consenso entre todos los segmentos de la sociedad británica y entre todos los matices de la opinión pública, de que la salud no debiera distribuirse de acuerdo con la capacidad para pagar, en otras palabras, «el comunismo en la salubridad».

Ni Blaug ni ningún otro analista competente negaría que hay graves problemas de asignación dentro del Servicio Nacional de Salud. Pero mientras subsista el consenso acerca del uso de los mecanismos de elección colectiva, habrá que encontrar soluciones al nivel de organización y administración sin comprometer la integridad del núcleo de la política. La heurística negativa del sistema prohíbe un retorno al principio de la

capacidad para pagar (por ejemplo, por medio de cargos generalizados al usuario), cualesquiera que sean los méritos de semejante opción en cuanto a su eficiencia en la asignación.

Desde luego, los servicios nacionales de la salud y los sistemas de seguridad social son casi únicos en sus dimensiones y su alcance, así como en lo explícito de sus compromisos ideológicos. Estas características favorecen la articulación de doctrinas programáticas generales, facilitando así la distinción entre principios nucleares y ajustes tácticos. Pero todas las políticas aceptadas mostrarán una estructura que, aunque tal vez menos bien articulada, no será cualitativamente distinta de los ejemplos que hemos analizado. Ciertamente es que los encargados de las políticas a menudo se desconciertan cuando se les exige una declaración precisa de los principios que guían sus acciones. Sin embargo, este estado de cosas es común en todo tipo de principios. Artesanos, hombres de negocios y atletas, con frecuencia saben lo que deben hacer en determinada situación, sin poder expresar en palabras los principios según los cuales actúan; esto no significa que sus acciones no procedan de acuerdo con unos principios. Y lo mismo puede decirse de los elaboradores de las políticas. Su incapacidad o su renuencia a enumerar las normas y los compromisos básicos no es prueba de que sus medidas políticas carezcan de un núcleo más o menos bien definido.

Evaluación de políticas

La metodología de los programas de investigación ofrece no solo una explicación de la continuidad y el cambio en la ciencia, sino también un conjunto de conceptos normativos para evaluar el desarrollo científico. Los programas rivales de investigación se evalúan en función de los desplazamientos de problemas que inducen (los problemas científicos, como los problemas de políticas, nunca se resuelven, sino que tan solo se desplazan). Un cambio de problema es progresivo si tiene un mayor poder explicativo y predictivo que las formulaciones anteriores. Un programa de investigación estará progresando si genera desplazamientos progresivos de problema; de otra manera, estará estancado.

Estos criterios sugieren algunas analogías potencialmente útiles para comparar programas de acción en competencia, como las políticas

keynesiana y monetarista. Sin embargo, se debe reconocer que una comparación directa de los programas de acción es extremadamente difícil por muchas razones, que incluyen la multiplicidad de normas evaluativas legítimas, la presencia de muchas causas concomitantes de éxito y de fracaso y la consiguiente ambigüedad de los registros históricos. Empero, la metodología ofrece algunas lecciones y advertencias de uso más inmediato a los evaluadores.

Lakatos subraya el punto de que *ponemos a prueba* una teoría o hipótesis aislada, pero *evaluamos* un programa de investigación. Dado que las teorías acuden a nosotros, no una a la vez, sino unidas en una red más o menos integrada de ideas, la unidad apropiada de evaluación no es una teoría aislada, sino todo un programa de investigación. A mi parecer, esta distinción es igualmente pertinente en el contexto de la evaluación de programas o de política. Por lo general, los evaluadores intentan poner a prueba los programas individuales, y no evaluar políticas enteras. A la luz de la dificultad de evaluar las políticas, esta es una restricción sensata. Sin embargo, a menudo los propios evaluadores se ven obligados a tomar en consideración el contexto más amplio de la política, sin comprender que puedan necesitarse diferentes consideraciones y normas evaluativas.

Los evaluadores, como buenos positivistas, se consideran científicos objetivos que ofrecen al administrador los hechos «duros», los «potenciales falsadores» del programa. En realidad, formulan metas, les asignan pesos relativos (que dependen de objetivos más amplios de la política), identifican actores e interesados, definen los límites de sistemas (que pueden incluir otros elementos del programa de acción) y escogen los raseros para medir. Parafraseando a Lakatos podríamos decir que el blanco de la flecha de la evaluación cobra forma cuando la flecha ya va en el aire. La evaluación no presupone un programa plenamente articulado; la evaluación lo recrea y, al hacerlo, lo relaciona con aquella política de la cual el programa no es más que un elemento.

Una vez más, los evaluadores (muy comprensiblemente, dadas las expectativas de sus clientes), ya por tradición se han preocupado por el alcance de los objetivos. Pero la esperanza de poder justificar o «falsear» un programa individual simplemente comparando los resultados con las

metas se desvaneció con el descubrimiento del ubicuo fenómeno del «poco efecto». Como escribe Weiss (1975), uno de los principales obstáculos al empleo de los resultados de una evaluación es precisamente su desalentadora tendencia a mostrar que el programa ha surtido poco efecto. Y no les va mejor a las organizaciones: «Medidas contra las olímpicas alturas de su meta, casi todas las organizaciones obtienen la misma calificación muy baja en eficacia. Las diferencias entre organizaciones tienen poca importancia» (Etzioni, 1960). Las teorías científicas, los programas y las organizaciones parecen nacer ya refutados, y la evaluación —como generalmente se la concibe y se la practica— no puede desempeñar un papel crucial en su desarrollo. No se debe confundir *refutación con rechazo*.

El fenómeno del poco efecto se vuelve menos sorprendente cuando reconocemos que una evaluación hecha exclusivamente en términos de resultados tiene que ser inconclusiva en circunstancias normales. Para seguir adelante con su trabajo, los evaluadores deben suponer que sus modelos y técnicas de medición no tienen problemas, o que son menos problemáticos que las hipótesis de trabajo incorporadas al programa que están evaluando. De hecho, una conclusión de poco o ningún efecto podría ser interpretada, hablando en abstracto, como una falla del modelo del evaluador, tan probablemente como una falla del programa.

Para suprimir la ambigüedad, puede ser necesario determinar cuál de las dos teorías o modelos embona mejor con las hipótesis y metas de la política de nivel superior. Pero esto muestra que no siempre y ni siquiera habitualmente es posible poner a prueba un programa particular sin evaluar o al menos interpretar la política de la que forma parte. En particular, es importante apreciar la posición del programa en relación con el núcleo de la política, ya que los programas cercanos al núcleo comparten algo de su inmunidad ante la crítica. Como lo muestra la anterior cita de Blaug (1970), las censuras de los analistas y evaluadores de las políticas a menudo parecen impertinentes no porque sean infundadas, sino porque van dirigidas a un blanco que no es el debido.

El «Mundo 3» de Popper y el desarrollo de la política

Como Lakatos, también Popper en sus últimas épocas se concentra en el crecimiento y desarrollo de las teorías científicas, más que en su refutación. Las teorías se desarrollan y crecen en un espacio cuasiautónomo de constructos intelectuales y de ideas en el sentido objetivo. Popper distingue tres «mundos» o niveles de realidad: primero, el mundo de los objetos físicos y estados físicos; segundo, el mundo de los estados mentales, de las preferencias subjetivas y de las creencias; tercero, el mundo de las estructuras objetivas, producidas por mentes humanas, pero que, una vez producidas, existen independientemente de ellas. Este es el mundo de las ideas en el sentido objetivo, de las teorías en sí mismas y de sus relaciones lógicas, de los argumentos en sí mismos y de situaciones problemáticas en sí mismas (véase Popper, 1972a, 1972b).

Este «Mundo 3» es autónomo de los otros dos niveles de realidad (precisamente en el sentido en que podemos decir que una obra de arte, una vez creada, cobra vida propia), aunque se relacione con ellos por cierto número de nexos y retroalimentaciones. Los tres mundos están relacionados de tal modo que el primero y el segundo pueden interactuar y los últimos dos también pueden interactuar. De este modo, el segundo mundo de experiencias y actividades subjetivas media entre los mundos primero y tercero. Las aplicaciones tecnológicas de una teoría científica pueden ejercer un impacto enorme sobre el mundo físico, pero solo a través del medio de las mentes y las decisiones humanas. Mas la relativa autonomía del Mundo 3 se revela por el hecho de que casi todo su crecimiento es consecuencia, a menudo imprevista, de teorías previas y de situaciones problemáticas que genera (por ejemplo, el concepto del quebrado aritmético conduce inevitablemente al problema de las magnitudes conmensurables e inconmensurables, y luego al descubrimiento de que la diagonal de un cuadrado es inconmensurable con su lado). La metodología de programas de investigación de Lakatos representa una explicación muy sutil del proceso de desarrollo de teorías en el Mundo 3, y añade criterios normativos para distinguir entre los desarrollos positivos y negativos.

Es importante comprender que el Mundo 3 de Popper incluye no solo las teorías científicas, sino también todos los productos de la mente humana vistos como constructos objetivos y no como simples expresiones subjetivas: idioma, normas éticas, leyes e instituciones. A esta lista podríamos añadir ahora las políticas, ya que también las políticas, una vez iniciadas, cobran vida propia. Interactúan entre sí, producen consecuencias imprevistas, crean problemas nuevos y quedan encarnadas en leyes, instituciones, normas y prácticas sociales. A veces, unas medidas que parecían producidas por el intento de un formulador de políticas por alcanzar una meta determinada muestran ser resultado de procesos políticos o burocráticos que de ninguna manera tenían en mente esa meta (Allison, 1971). O, como en Vietnam, los responsables de las políticas dan un paso y el siguiente se sigue de él. En resumen, el desarrollo de políticas muestra unos rasgos objetivos que no pueden interpretarse, sencillamente, diciendo que los formuladores de las políticas cambiaron de opinión.

Para comprender estos rasgos objetivos necesitamos algo similar a la distinción que establece Popper entre tres niveles de realidad. Donde Popper habla de tres diferentes «mundos», yo prefiero hablar de tres diferentes «espacios»: el espacio de políticas (*policy space*), el *espacio del actor* y el *espacio del problema*. El espacio de políticas es un conjunto de políticas interrelacionadas (o programas de acción) disgregadas en sus elementos nucleares y cinturones protectores. Decir que las políticas están interrelacionadas significa que sus cinturones protectores intersecan en algunos puntos. En otras palabras, el mismo programa o propuesta de la política lleva diferentes «caras», según el núcleo desde el cual se lo ve. Así, para tomar un ejemplo propuesto por Allison (1971: 168) y ponerlo en mi terminología, una propuesta de retirar de Europa tropas norteamericanas es para el programa de acción del ejército de los Estados Unidos una amenaza a su presupuesto y su tamaño; para el programa de acción de la Tesorería es una ganancia de su balanza de pagos, y para el Departamento de Estado es una amenaza a las buenas relaciones con sus aliados de la OTAN. Un espacio de políticas puede analizarse por sus características estructurales genéricas, como:

- 1) *grado de fragmentación* (por ejemplo, transportes, energía o educación) o *concentración* (por ejemplo, seguridad social, salud);

- 2) *nivel de madurez* (áreas de política nuevas contra antiguas; por ejemplo, protección del ambiente y administración de riesgos contra impuestos o política exterior);
- 3) *nivel de competencia* de las fuentes privadas o públicas (por ejemplo, defensa contra salud o educación);
- 4) *exposición a influencias internacionales* (políticas industriales o agrícolas, contra seguridad social o educación).

Como lo ha sugerido la frase de Lowi (1964), «las políticas determinan la política», las características estructurales del espacio de políticas afectan la estructura del espacio del actor, por ejemplo, los tipos de coaliciones que se pueden formar, o el grado de fragmentación de la «comunidad de políticas» (Kingdon, 1984). Una comunidad de políticas se compone de especialistas que comparten un interés activo en una determinada área de política, por ejemplo: académicos, analistas, planificadores de las políticas, miembros del personal congresional, expertos sindicales o de grupos de intereses y profesionales. Los miembros de una comunidad de políticas representan diferentes intereses, sostienen distintos valores y pueden estar dedicados a diversos programas de investigación, pero todos ellos contribuyen al desarrollo de las políticas, al introducir y discutir nuevas ideas y propuestas de políticas.

Aunque algunos miembros de una comunidad de políticas también puedan ser actores políticos, sus dos roles son totalmente distintos. Un votante que elige en un referendo o un formulador de política que elige entre diferentes opciones no contribuyen a la innovación conceptual; antes bien, actúan como mecanismos para seleccionar del portafolio de variantes de política disponibles. De hecho, se puede considerar al proceso político como el mecanismo de selección que elige, para su aceptación, de entre las propuestas de política en competencia la que mejor satisface las demandas del medio político.

En esta perspectiva evolutiva, la innovación de políticas se analiza como resultado de un proceso doble (que ocurre en el espacio del actor) de variación conceptual y selección política. La comunidad de políticas es el

sitio de la innovación conceptual, en tanto que la arena política es el sitio de la selección. Las comunidades de políticas varían mucho en su riqueza de conexiones y en otras características estructurales, como centralización o fragmentación, intensidad de la competencia y tasa de innovación conceptual. Estas diferencias dependen, en parte, de las características del espacio de políticas (como ya se observó), pero también de las características de la arena política y de los programas de investigación de las profesiones y disciplinas académicas pertinentes.

Por ejemplo, hasta hace poco, los críticos expertos del sistema de seguridad social (en los Estados Unidos y en Europa) carecieron de una base organizacional por la cual tuvieran acceso al foro de políticas correspondiente. Como lo muestra Derthick (1979), la profesión actuarial en los Estados Unidos había producido algunos penetrantes análisis de los problemas de la seguridad social a finales de los treinta, pero después del periodo inicial rara vez hubo discusiones críticas. Las convenciones actuariales en general dejaron de prestar gran atención al tema. De manera similar, la profesión de la economía solo durante los sesenta empezó a desarrollar una cantidad sustancial de análisis fundamentales de la seguridad social. Los economistas que escribían sobre seguridad social antes de esa fecha se encontraban fuera de la corriente principal de su disciplina (como Eveline Burns, quien pasó la mayor parte de su vida profesional enseñando en la Escuela de Trabajo Social de la Universidad de Columbia), o sus intereses iban dirigidos en realidad a cuestiones más generales y distintas, por ejemplo: los economistas keynesianos, quienes consideraban la seguridad social básicamente como un instrumento de política fiscal. Los ejecutivos del programa de seguridad social y sus partidarios en el Congreso fácilmente podían pasar por alto las críticas que les llegaran de expertos, carentes de un difundido apoyo profesional (Derthick, 1979: 158-182). Todo esto cambió con las recurrentes crisis financieras de años recientes. Estas crisis y un cambiante clima de la opinión acerca del papel apropiado del gobierno en la economía han estimulado un esfuerzo intelectual mucho más sostenido de los economistas y de otros analistas, para desarrollar propuestas más o menos radicales tendientes a reformar el sistema actual.

Como lo muestra este ejemplo, la falta de acceso a un foro apropiado puede ser, en sí misma, un obstáculo total a la consideración debida de nuevas ideas de política. El desarrollo y la evaluación sistemática de nuevas propuestas son imposibles sin oportunidades bien organizadas de debate crítico. La comunidad de políticas debe estar lo bastante abierta y competitiva para que puedan surgir variantes verdaderamente nuevas. El empleo de fuentes externas de conocimiento, asesoría y evaluación por las agencias del gobierno es una manera de aumentar la corriente de las innovaciones conceptuales. Al mismo tiempo, la selección solo puede ser eficaz cuando la comunidad no está demasiado abierta. Si se tomara en serio cada propuesta, pronto se haría intolerable la carga para los mecanismos de selección, conduciendo a un desplome de las normas de evaluación. Un rasgo característico de una comunidad intelectual en buen estado de funcionamiento es que cada argumento presentado debe relacionarse con algún punto de vista ya presente en la comunidad. Una nueva propuesta será juzgada no solo por sus méritos intelectuales, sino también por el modo en que pueda contribuir al debate en marcha y relacionarse con los actores que haya en el campo intelectual y con sus intereses (Unger, 1975: 110-115). Por tanto, solo algunas de las propuestas y variantes políticas teóricamente concebibles pueden volverse temas activos de debate y de innovación en un momento dado.

Además de evaluar las políticas en vigor y discutir sobre las nuevas propuestas de políticas, los miembros de una comunidad de políticas supervisan constantemente el espacio del problema, con objeto de identificar los problemas nacientes y detectar los cambios estructurales que surjan en la economía y en la sociedad. Al explorar el espacio del problema, estos especialistas no se dejan guiar tanto por sus intereses académicos o intelectuales como por el deseo de descubrir puntos nuevos que sean prometedores para la agenda de las políticas.

Como el espacio de las políticas y el del actor, también el espacio del problema puede describirse de acuerdo con sus características estructurales genéricas o específicas. Estas características determinan, de manera importante, el modo en que la comunidad de políticas y los actores políticos perciben y tratan diferentes cuestiones. Un temprano ejemplo sobresaliente de análisis estructural de un espacio de problemas

puede encontrarse en *Journeys Toward Progress* (1963), de Hirschman, en que el autor contrasta los rasgos característicos del problema de inflación estudiado en Chile con los del subdesarrollo regional (en Brasil) y el mal reparto de las tierras (en Colombia). Afirma que la inflación, en contraste con el mal reparto de las tierras, emite señales regulares, puede medirse de un mes a otro y, por ello, es susceptible de recibir atención especial de analistas y responsables de políticas en cuanto pase algún umbral de tolerancia. Por otra parte, en contraste con las calamidades naturales como las sequías periódicas que afligen al noreste de Brasil, la inflación es considerada como una calamidad debida a la mano del hombre y que, por tanto, se presta a recriminaciones entre los grupos sociales. A diferencia del desarrollo regional, la inflación es bastante similar entre un país y otro. También se trata del tema sumamente técnico del dinero. Vemos así que las doctrinas dominantes y los expertos internacionales desempeñan papeles de particular importancia en los intentos de solución, aunque tanto expertos como doctrinas también puedan desempeñar, en alto grado, papeles políticos. Una vez más, la inflación es fácil de «sobrerresolver»; súbitamente, puede dejar el paso a la deflación. Por tanto, el temor a la deflación puede inhibir a los formuladores de las políticas que se enfrentaban a la inflación, mientras que la elaboración de las políticas sobre el subdesarrollo regional, o sobre un mal reparto de tierras, puede proceder desde considerable distancia sin peligro de llegar demasiado lejos (Hirschman, 1963, 1981).

A su vez, esas características específicas pueden subsumirse en categorías más generales, como «problemas de política, así como problemas de política privilegiados y desdeñados». Descubrir esas características estructurales es parte importante de la labor del analista; por medio de ellas, es posible comprender y tal vez predecir vías alternas de hechura de las políticas. Sin embargo, un análisis detallado de estos temas superaría, con mucho, el limitado espacio de nuestro capítulo.

Conclusiones

He tratado de mostrar, con la ayuda de argumentos y ejemplos, que algunos conceptos desarrollados por los filósofos de la ciencia durante los últimos veinte años pueden adaptarse a las necesidades de los estudiosos

de la política pública. Como modelo normativo, la metodología de los programas de investigación subraya la importancia de determinar la unidad correcta de evaluación, y señala cierto número de trampas potenciales contra las que deben estar en guardia todos los evaluadores. La misma metodología, empleada como modelo descriptivo en relación con el análisis de la autonomía del Mundo 3 de Popper, sugiere maneras interesantes de analizar la estructura sutil de las políticas y su desarrollo con el transcurso del tiempo.

Podrá decirse que los conceptos introducidos en este capítulo solo ofrecen un nuevo vocabulario para ideas que los analistas de las políticas siempre han utilizado, al menos implícitamente. Aun si fuera así, valdría la pena subrayar una vez más la importancia metodológica de emplear conceptos bien articulados al estudiar fenómenos tan complejos como los de la hechura de las políticas. En realidad, creo que algunas de las ideas aquí presentadas pueden ser útiles para arrojar nueva luz sobre algunos de los problemas más importantes del análisis de políticas.

Por ejemplo, la distinción entre el núcleo y el cinturón protector no solo expresa la idea de que la continuidad y el cambio son inherentes al desarrollo de las políticas, sino también explica por qué el incrementalismo es un rasgo tan continuo de todos los sistemas de elaboración de políticas. Además, no todos los pasos incrementales de un tamaño determinado tienen la misma importancia, pues la significación del cambio incremental está en relación inversa con su distancia del núcleo de políticas. Por tanto, los partidarios del cambio no incremental harían bien en estudiar minuciosamente la topología del espacio de políticas, en el que estén planeando concentrar sus esfuerzos.

Asimismo, el concepto de una heurística positiva y negativa (el hágase y el no se haga de un núcleo de política) parece ser una manera prometedora de reformular el concepto de la no toma de decisiones en términos operativos. Acaso sea verdad, como señalan sus críticos, que una no decisión es un no acontecimiento y que, por tanto, no se la puede estudiar empíricamente. Pero una heurística negativa, como su gemela la heurística positiva, es un auténtico sujeto de estrategia para la observación y el escrutinio crítico.

Por último, sostendré que uno de los problemas más desconcertantes para los analistas de las políticas y los estudiosos de la estrategia de negocios —el surgimiento de políticas o estrategias que no se había propuesto ninguna persona, grupo ni organización— solo puede comprenderse como un fenómeno del espacio de políticas, como la consecuencia imprevista de procesos que ahí tienen lugar. Al menos desde el tiempo de David Hume, el estudio de las consecuencias imprevistas o los productos no planeados de las acciones humanas se ha considerado como la principal tarea de la ciencia social. El Mundo 3 de Popper (del cual es un subconjunto el espacio de políticas) aporta el ya muy necesario fundamento epistemológico de tal estudio.

Notas

* Título original “Social Sciences and Modern States”. Primera edición en inglés ©1991 por University of Cambridge Press. Traducción autorizada al español “Programas de investigación y programas de acción o ¿puede la investigación de políticas aprender de la ciencia?”, en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.), “Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas”, pp. 327-356, Fondo de Cultura Económica, México, 1999. Reproducción autorizada.

Referencias

- Allison, Graham T., 1971. *Essence of Decision*. Boston, Little, Brown, and Company. [Traducido al español por Grupo Editor Latinoamericano].
- Blaug, Mark, 1970. *An Introduction to the Economics of Education*. Harmondsworth, Middx. Penguin Books.
- , 1975. «Kuhn versus Lakatos, o Paradigms versus Research Programs in the History of Economics». *Hope* 7, pp. 399-433.
- Boothroyd, Hylton, 1978. *Articulate Intervention*. Londres, Taylor and Francis.
- Derthick, Martha, 1979. *Policymaking for Social Security*. Washington, DC, The Brookings Institution.
- Etzioni, Amitai, 1960. «Two Approaches to Organizational Analysis: A Critique and a Suggestion». *Administrative Science Quarterly* 5 (abril-junio), pp. 257-278.
- Hecló, Hug, 1972. «Review Article: Policy Analysis». *British Journal of Political Science* 2 (enero), pp. 83-108.

Hirschman, Albert O., 1963. *Journeys toward Progress: Studies of Economic Policy Making in Latin America*. Nueva York, The Twentieth Century Fund.

——, 1981. *Essays in Trespassing*. Cambridge University Press.

Kingdon, John W., 1984. *Agendas, Alternatives, and Public Policies*. Boston, Little, Brown and Company.

Kuhn, Thomas S., 1970. *The Structure of Scientific Revolutions*. 2a. ed., University of Chicago Press. [Hay traducción al español en el FCE].

Lakatos, Imre, 1968. «Criticism and the Methodology of Scientific Research Programmes». *Proceedings of the Aristotelian Society* 69, pp. 146-186.

——, 1971. «History of Science and Its Rational Reconstruction». En *Boston Studies in Philosophy of Science*. Ed. R. S. Cohen y C. R. Bucks. [Hay traducción al español en Ian Hacking (comp.), *Revoluciones científicas*, México, FCE, 1985.]

Lippmann, Walter, 1943. *U.S. Foreign Policy: Shield of the Republic*. Boston, Little, Brown and Company.

Lowi, Theodore, 1964. «American Business, Public Policy, Case Studies, and Political Theory». *World Politics* (julio), pp. 677-715. [Hay traducción al español en Luis F. Aguilar, *La hechura de las políticas*, México, Miguel Ángel Porrúa, 1992].

Popper, Karl R., 1972a. «Epistemology without a Knowing Subject». *Objective Knowledge*. Londres, Oxford University Press.

——, 1972b. «On the Theory of the Objective Mind». *Objective Knowledge*. Londres, Oxford University Press.

Unger, Roberto Mangabeira, 1975. *Knowledge and Politics*. Nueva York, Free Press. [Hay traducción al español en el FCE].

Weiss, Carol, 1975. «Evaluation Research in the Political Context». *Handbook of Evaluation Research*, pp. 13-25. Vol. 2. Ed. E. L. Struening y M. Guttentag. Londres, Sage Publications.

La investigación de políticas: ¿datos, ideas o argumentos?*

Carol H. Weiss

Hemos aplazado ya bastante el examen de los efectos de la ciencia social sobre la política pública. Este capítulo y el siguiente, obra de Wittrock, enfocan ahora los modos en que las ciencias sociales influyen en el desarrollo de las políticas en el Estado moderno. Mientras Majone (capítulo XIII) tomó ideas de la filosofía de la ciencia, yo adopto una idea tomada del sistema jurídico: la idea de argumentación.

En la primera mitad del capítulo, examinaré la influencia de tres tipos de productos de la investigación sobre las políticas: datos y descubrimientos, ideas y crítica, y argumentos o informes para la acción de políticas. Mientras que el producto tradicional de un estudio de políticas es un informe del primer tipo, lleno de datos, conclusiones, estadísticas y descubrimientos, una revisión de los someros testimonios disponibles parece indicar que, en ciertos medios, la investigación ejerce mayor impacto cuando pasa a ser parte de la defensa activa de una posición preferida.

Así, la segunda parte del capítulo busca una respuesta a la pregunta normativa: ¿qué actitud deben adoptar los investigadores? Se enfrenta a la pregunta de si la defensa activa, la promoción y la recomendación tienen un lugar en la agenda del investigador de las políticas. La investigación de políticas es parienta cercana de la ciencia social, y aunque se haya puesto sus ropas de trabajo y se haya ido a laborar en las oficinas y antecámaras del gobierno, no ha abandonado su nombre de «ciencia»: de ahí las «ciencias

de las políticas». Hay algo incómodo en pensar siquiera en abandonar las normas de objetividad que caracterizan una ciencia y abrazar en cambio la noción de la defensa activa, la promoción y la recomendación, más apropiada para un grupo de interés o de cabilderos. Y sin embargo, los avances ocurridos en la epistemología (y hasta en la mecánica cuántica) han socavado los conceptos de una sola realidad «que está ahí». En el periodo pospositivista, todos tenemos clara conciencia de que los investigadores «construyen» el mundo que estudian y de que los valores, prioridades y modelos conceptuales que ponen en su trabajo influyen sobre las cosas que ellos «descubren». Si la objetividad es en realidad una ilusión, ¿es la recomendación, tal vez, un franco reconocimiento del carácter parcial de su obra? Yo batallo con este dilema e intento establecer las implicaciones de los diversos cursos de acción para los investigadores de políticas y para las instituciones para las que trabajan y a las que sirven.

Las cuestiones que enfocamos en este capítulo son importantes en varios niveles. La cuestión empírica —en qué forma ejerce mayor efecto la investigación de políticas— tiene un enfoque aplicado; a aquellos científicos sociales que intentan aumentar la influencia del análisis empírico sistemático sobre la elaboración de las políticas públicas les convendría saber en qué forma de presentación es más probable ejercer un impacto en diferentes arenas. Si supiéramos lo suficiente, podríamos ayudar a hacer que sus informes tuviesen el máximo efecto.

La cuestión del proceso —cómo la información interactúa con otros elementos en la formulación de las decisiones públicas— posee importantes elementos sustantivos. Tiene implicaciones para el papel de los expertos —en este caso, los investigadores de las ciencias sociales— en el sistema de elaboración de las políticas. Si los investigadores logran influir más sobre las políticas cuando van más allá de su papel tradicional de aportar evidencia empírica, entonces encontraremos preguntas acerca de los efectos probables de su ingreso más dinámico en los conciliábulos de la hechura de políticas. En particular, si utilizan la investigación para hacer avanzar valores de políticas, ¿pondrán en peligro su «boleto de entrada», el cual se basa en suposiciones de un conocimiento objetivo y que se impone? ¿Se enajenarán a algunas partes del espectro político? ¿Servirá su recomendación para dar una ventaja especial a los intereses

que favorecen? ¿Pondrá en peligro los intereses de la ciencia social institucional? ¿O resistirán los científicos sociales las presiones que se les hagan para contratarlos como activistas argumentativos, en favor de unas orientaciones de política particulares o de políticas específicas? A la postre, tendremos que enfrentarnos a la pregunta normativa de cuál es el papel apropiado para un experto en una entidad política de carácter democrática.

Desde el principio, reconozcamos que la investigación de políticas o cualquier otro tipo de investigación no va a determinar la principal dirección de la política. Políticos y funcionarios poseen convicciones ideológicas y constelaciones de intereses, que en gran parte determinan el rumbo que ellos siguen. El lugar de la información en general y de la información de la investigación en particular deben verse, antes bien, en el sentido de ayudar a los elaboradores de políticas a decidir cuáles medidas son las más apropiadas para la realización de sus ideologías y sus intereses (Weiss, 1983). No se espera que los investigadores participen como tomadores de decisiones. En la esfera de la política pública, su tarea ha sido, en general, iluminar las consecuencias de las diversas opciones para que quienes ocupan los cargos de autoridad puedan saber lo que obtendrán y aquello a lo que renunciarán al seleccionar un rumbo particular. La investigación de políticas es un actor de reparto en el drama de la hechura de políticas. La pregunta empírica es ¿en qué condiciones entra siquiera la investigación en escena y, cuando lo hace, qué consecuencias tendrá para el desenvolvimiento de la acción? Aunque creo que esta generalización es cierta en países del mundo entero, es importante observar que casi toda la evidencia de este capítulo se deriva de la experiencia norteamericana.

La investigación de políticas como datos

Hasta hace cerca de un decenio se daba por sentado que lo que deseaban los formuladores de las políticas era precisamente lo que los investigadores eran más capaces de ofrecer: o sea, datos, descubrimientos y conclusiones de investigación. Los científicos sociales presuponían que los funcionarios políticos y burocráticos encontrarían el proceso de toma de decisiones considerablemente simplificado si supieran, por ejemplo, cuántas personas que viven en la pobreza eran adultos, sin impedimentos físicos y sin niños pequeños; si la enseñanza para niños con un limitado dominio del inglés

ha sido mejor si se les enseña en la lengua materna de los niños o en inglés; si una reducción del impuesto a las ganancias del capital aceleraría el crecimiento económico o reduciría el ingreso del gobierno. Sin duda, las políticas podrían beneficiarse si contaran con mejores datos sobre la situación de las condiciones actuales, la evaluación de la eficacia de los programas pasados y actuales y las probables consecuencias futuras de las diversas opciones de acción.

Todo esto parecía muy razonable. Lo malo era que cuando los científicos sociales se ponían a estudiar los efectos de la investigación sobre las decisiones del gobierno, encontraban muy poco efecto (Caplan, 1977; Bulmer, 1978; Weiss y Bucuvalas, 1980; Alkin *et al.*, 1979; Deitchman, 1976; Dockrell, 1982; Knorr, 1977; Rich, 1977; Leff, 1985; Walker, 1987). Aun cuando se encargaran explícitamente estudios de políticas para responder a preguntas específicas del gobierno, no parecían tener gran efecto sobre las decisiones ulteriores. Los estudios de evaluación fueron un ejemplo notable. En muchos ámbitos, se encargaron evaluaciones para descubrir hasta qué punto programas generosamente patrocinados por el gobierno estaban alcanzando sus metas con objeto de tomar decisiones acerca de su continuación, terminación o cambio. Pero los resultados de la evaluación parecieron ejercer muy poca diferencia sobre el destino de los programas estudiados (véase también Majone, capítulo XIII). El programa *Head Start* salió casi incólume de una temprana evaluación desfavorable y luego procedió a florecer. Los *Job Corps* se enfrentaron a evaluaciones de costo-beneficio que habrían debido conducir a su temprana muerte, pero conservaron su apoyo político. Por otra parte, la favorable evaluación de las becas de *college*, de acuerdo con la Ley Nacional de Educación para la Defensa, no bastó para salvar al programa de severos recortes y de su terminación.

Lo que ocurría era que otros valores e intereses tomaban precedencia sobre los datos. Si los encargados de las políticas deseaban mostrar a sus votantes que estaban «haciendo algo» contra el desempleo, para ellos no significaba mucha diferencia que las políticas de preparación en el empleo que ellos habían apoyado estuviesen ejerciendo un impacto mínimo. Al menos, ahí quedaban registradas, diciendo que apoyaban el empleo. Si otro grupo de formuladores de políticas estaba comprometido, por principio,

con la desregulación y la privatización, no estaba interesado en escuchar los descubrimientos acerca de la ocasional superioridad de los servicios suministrados públicamente. Esa opción estaba fuera de lo que podían aceptar. Los descubrimientos de la investigación se utilizaban básicamente cuando ayudaban a los actores de las políticas a hacer lo que ya deseaban hacer. En tales circunstancias, los datos les ayudaban a ajustar, adaptar y afinar sus propuestas para que embonaran mejor en los hechos del caso.

Algunos datos se toman en cuenta ya por rutina en la toma de decisiones gubernamental. Ciertos tipos de datos económicos parecen convenir particularmente, porque los datos económicos ocupan un lugar central en las decisiones acerca de impuestos y gastos. Quiéranlo o no, los gobiernos se ven envueltos en la política macroeconómica y llegan a depender marcadamente de las previsiones económicas. Al enfrentarse a cuestiones específicas como la solvencia del Fideicomiso de Seguridad Social, los datos económicos ejercieron gran influencia. Pero a veces hasta los datos económicos se utilizan más como «munitiones» para apoyar posiciones de políticas ya existentes, que como señales de guía. Los mismos datos han aparecido en ambos lados de una cuestión, utilizados en apoyo de prescripciones radicalmente distintas... de acuerdo con las inclinaciones políticas de sus partidarios.

Algunos datos estadísticos y de investigación de varias índoles sí han afectado los contornos de las políticas, como los datos sobre las necesidades de salud de los ancianos (que influyeron sobre la aprobación de *Medicare*), las tasas de ocupación de camas de hospital (que influyeron en las decisiones sobre la construcción de hospitales), la prevalencia del analfabetismo de adultos (el estudio efectuado por el National Assessment of Educational Progress mostró tasas inferiores a las esperadas y redujo la importancia política de esta cuestión en 1986) y las tasas de desempleo (utilizadas en apoyo de toda una variedad de programas de capacitación laboral y creación de empleos). Algunas fórmulas del gobierno, como las utilizadas para calcular los gastos y reembolsos de *Medicare*, la distribución de la ayuda federal a los estados, los planes de retiro del servicio civil, etc., dependen de los datos; los que toman las decisiones se ven obligados a prestar atención a las cifras.

¿Quién es probable que transmita la investigación de políticas en forma de datos a los funcionarios del gobierno? Obviamente, los propios investigadores y analistas. Escriben informes a la manera en que los investigadores escriben sus informes, con datos, cuadros, gráficas y tal vez unas cuantas fórmulas en el apéndice, junto con su interpretación de lo que significan los datos. Una concesión habitual a la esfera de las políticas es que incluyen un «sumario ejecutivo», en que resumen los puntos clave en unas cuantas páginas y a veces tratan de sacar las implicaciones de los datos para las opciones de políticas a que se enfrentarán quienes tomen las decisiones. El sumario ejecutivo suele estar bien escrito, pero la aplicación de los descubrimientos a las actuales opciones de políticas a veces no da en el blanco; no todos los investigadores están lo bastante cerca del «juego interno» de la hechura de políticas para tener siquiera un atisbo de las opciones que se están considerando más seriamente; y para que sus esfuerzos sean oportunos y pertinentes, deberán revelar más que su condición de simples aficionados en la arena de las políticas.

Por encima de todo, los descubrimientos de la investigación no han tenido tanta influencia sobre las decisiones legislativas o burocráticas en materia de políticas como originalmente habían esperado los científicos sociales. A mediados de los setenta, cuando se puso en marcha el estudio empírico de la situación, los especialistas se sintieron perplejos y desalentados. Varias fueron las razones de su desencanto, y no fue la menor de ellas el altísimo nivel de las expectativas que habían puesto en el tema. Consideraban la toma de decisiones como una empresa racional y esperaban que la información ejerciera una influencia directa. Muchos de ellos creían en el antiguo adagio de que saber es poder, sin comprender que en la política suele ocurrir, más a menudo, lo contrario: poder es conocer... o al menos da acceso y control del conocimiento. Los científicos sociales también fueron ingenuos al no comprender la tenacidad de las convicciones ideológicas y del interés egoísta de las organizaciones. De hecho, su optimismo original y desencanto ulterior casi los cegaron ante los cambios muy reales, pero sutiles y lentos, que causó la investigación de las políticas.

La investigación de políticas como ideas

Lo que encontramos a mediados y finales de los setenta fue que la investigación de políticas servía menos de dirección a las políticas, que de «ilustración» general para los elaboradores de las políticas. Estos no estaban utilizando tanto datos sueltos cuantitativos o cualitativos, o conjuntos de descubrimientos, sino eran influidos por las generalizaciones basadas en los datos.

Lo que distingue a la investigación de políticas como ideas es que se ha prescindido de la mayor parte de todo el atavío de la investigación. Se pierde lo específico de la metodología de investigación. Los encargados de las políticas y sus ayudantes no tienen la menor idea del carácter de la muestra, el número de casos, el contenido de las intervenciones que haya habido, las preguntas específicas planteadas, cómo se procesaron y analizaron los datos, ni de la calidad del estudio como investigación: cualquiera de los puntos específicos que les capacitarían a ver lo bueno que había sido el estudio, la credibilidad que debían darle y en qué situaciones sería razonable generalizar. Esta condición a menudo se sostiene aun cuando la investigación se presente como datos. La característica más reveladora de la investigación como ideas es que también han desaparecido los descubrimientos reales del estudio. Lo que la gente recuerda es que los programas de preparación en el trabajo funcionan para las madres que reciben pagos de la asistencia pública, pero no para los adolescentes o para los desempleados crónicos. Los descubrimientos originales de la investigación se reducen a una simple «historia», los enunciados condicionantes se pierden, y las conclusiones a menudo se extienden más allá de los descubrimientos del estudio. De este modo, los actores de la política recuerdan que la educación preescolar mejora los logros entre los niños menesterosos en la escuela, pero no recuerdan que el Programa Preescolar de Perry, que fue la fuente de muchos datos, tuvo un total de 58 niños en el grupo experimental, y que las diferencias en el cociente intelectual entre los experimentales y los de control desaparecieron a la edad de 14 años (Schweinhart y Weikart, 1980). Lo que queda es que una temprana intervención en la educación de los jóvenes es mejor que aguardar hasta que sea tarde. Entonces se aplica la idea (mucho más de lo que la justificarían los datos) para influir

sobre el desarrollo de los tempranos programas para la educación, salud y rehabilitación de la niñez.

En el aspecto positivo, las ideas tomadas de la investigación de las políticas pueden darnos nuevos atisbos del proceso de las políticas. Pueden alterar el modo en que la gente conceptualiza las cuestiones y enmarca los problemas (véase también Rein y Schon, capítulo XII). Las ideas pueden modificar el modo en que la gente percibe los elementos de una situación que pueden modificarse y los que tienen que aceptarse como dados. Pueden alterar la importancia de un problema y ayudar a darle mayor o menor importancia en la agenda pública. Las ideas pueden ayudar a determinar, por ejemplo, qué faceta de un problema deberá seleccionarse para la acción, si se debe gastar dinero en investigación médica o en tratamiento o prevención. Con el correr del tiempo, la asimilación de las ideas de la investigación puede ayudar a transformar el modo en que se enfocan las cuestiones de política y los tipos de solución que se consideran, se debaten y se adoptan.

¿Quién transmite la investigación de política en forma de ideas? Ante todo, los intermediarios de la investigación. Los especialistas han estado escribiendo durante más de 20 años acerca de los «intermediarios», en el proceso de propagación de la investigación (Lazarsfeld *et al.*, 1967; Wilson, 1978). Los escritos han sido caprichosos o bien exhortatorios, y piden la creación de una ocupación especializada que se encargue de convertir los descubrimientos de la investigación en prescripciones prácticas para quienes toman las decisiones. El subtexto de los escritos sobre los intermediarios ha sido que quienes adopten la especialidad de «hacer contactos» deben su lealtad a la investigación y no a los practicantes o elaboradores de las políticas (véase también Lippitt, 1969; Miles, 1964). Lo que rara vez se ha notado en las discusiones sobre los intermediarios de la investigación es que ya hay cierto número de personas encargadas de esos roles de vinculación; incluyen a asesores, preparadores, oradores, cabilderos, periodistas y otros (Weiss, 1986).

Los propios investigadores pueden ser intermediarios al transmitir el tema general de su labor o de la de otros. Cuando sirven como asesores de dependencias gubernamentales o como miembros de comisiones seleccionadas, su aportación distintiva es la de ideas desarrolladas por

medio de la investigación. El escritor profesional que elabora el «resumen ejecutivo» acerca de un estudio para su publicación, el periodista que escribe un reportaje sobre un estudio para la prensa semanal o cotidiana, el profesorado universitario que incluye los descubrimientos de la investigación en sus conferencias en el aula para educar a la siguiente generación, los preparadores, los participantes en las conferencias, quienes hablan en banquetes, desayunos y reuniones anuales, amigos y conocidos —todas las diversas personas que hayan oído los resultados de la investigación, en persona o de cuarta mano— las transmiten a quienes toman las decisiones. Algunos de ellos lo hacen intencionalmente, porque los descubrimientos de la investigación apoyan una actitud que ellos desean imponer, como suelen hacerlo los cabilderos; a veces actúan neutralmente, y esto es parte de su propio trabajo, como cuando los reporteros escriben noticias acerca de la investigación; a veces lo hacen ociosamente, como simple tema de conversación. En todas estas comunicaciones se pierden el tono y el detalle de la investigación original, y queda muy poca seriedad científica. Se hace hincapié en el meollo del relato. Por ejemplo, el tiempo que los estudiantes dedican a las tareas es lo que establece la diferencia acerca de cuánto aprenden; las terapias a largo plazo contra el alcoholismo no son más eficaces que las intervenciones breves. Estos son los tipos de ideas que «gotean» hasta la esfera de las políticas, desde toda una variedad de fuentes que ayudan a establecer el clima de la opinión para tomar decisiones ulteriores (Weiss, 1980, 1986).

Aunque hemos empleado el noble término de ilustración para describir la circulación de la investigación como ideas, desde luego no hay ninguna garantía de que las ideas sean correctas... o de que sean mejores que las que teníamos antes. La investigación de la que se derivaron pudo ser inadecuada, conceptual o metodológicamente; su abreviación en resúmenes o «lemas» puede haber omitido elementos vitales; la traducción pudo ser una completa deformación. Y sin embargo, las ideas procedentes de la investigación entran en circulación. Al ser absorbidas por el pensamiento convencional moldean las suposiciones de la gente acerca de lo que es importante, de lo que debe hacerse y de las soluciones que probablemente ayudarán a alcanzar los fines deseados.

Una serie de estudios que se han realizado desde 1975 sobre los efectos de la investigación mostraron que este es el «uso» más característico de la investigación (Aaron, 1978; Banting, 1979; Barber, 1987; Bulmer, 1978, 1982, 1987; Cherns, 1979; Derthick y Quirk, 1985; Feller, 1986; Hayes, 1982; Husén y Kogan, 1984; Kallen *et al.*, 1982; Patton *et al.*, 1977; Rich, 1981; Sabatier, 1978; Saxe, 1986; Weiss, 1977, 1980, 1986, 1989; Weiss y Bucuvalas, 1980). Mientras que la investigación como datos altera el curso de acción en esferas limitadas, muchos funcionarios dicen que las ideas llegadas de la investigación han alterado el modo mismo en que definen los problemas. Les parece difícil citar los efectos de la investigación sobre decisiones en concreto, pero tienen la sensación de que imbuje su comprensión y su labor.

Donde las ideas procedentes de la investigación tienen que ver con insuficiencias de una política actual provocan dudas acerca del *statu quo*. Refuerzan el escepticismo existente y sugieren que los responsables de las políticas deben reexaminar lo que se está haciendo y reconsiderar lo que se debe hacer. A menudo, favorecen una reevaluación fundamental de las políticas que están en vigor. Por ejemplo, las evaluaciones de los planes federales para contener el alza de los costos de hospitales, como los esfuerzos por reducir el gasto de capital y también la permanencia de los pacientes mostraron que las regulaciones no estaban funcionando. Al mismo tiempo, se descubrió que ciertos planes de reembolso habían surtido cierto efecto sobre la reducción de costos. Entonces, los investigadores crearon un sistema de «grupos relacionados con el diagnóstico» (DRG) para rembolsar a los hospitales de acuerdo con el diagnóstico del paciente; el sistema fue adoptado por el gobierno federal. Otro ejemplo fue la investigación que mostró que las innovaciones impuestas «de arriba abajo» a las escuelas solían durar muy poco. La idea estimuló a los funcionarios de la educación y a los legisladores, para canalizar fondos hacia innovaciones creadas localmente, en lugar de imponer cambios que todas las escuelas debieran acatar. La función de la investigación como crítica básica de las políticas puede, con el tiempo, tener efectos de largo alcance.

La investigación de políticas como argumentos

La investigación de políticas como argumentos representa no solo una investigación que se ha diluido, sino también una investigación a la que

se ha añadido algo. Lo que se ha añadido es una actitud de defensa activa y recomendación. Cuando esto ocurre, no solo se pierden algunos de los datos, como en la investigación como ideas, sino también se pierden selectivamente los datos. Los descubrimientos que favorecen «la otra mano» o que debilitan el poder del argumento se descartan, con objeto de hacer más convincente el argumento.

Desde luego, gran parte de la investigación de políticas se ha planeado para considerar los efectos pasados o los potenciales futuros de distintas políticas, y sus descubrimientos casi siempre preferirán un curso sobre otro. Lo que tengo en mente no es esa preferencia con base empírica; antes bien, la investigación como argumentos comienza con un conjunto de valores (sean los del investigador, los del cliente o los del empleador), y solo considera opciones dentro de esta gama ideológica. La restricción de la perspectiva puede ser involuntaria, como cuando los investigadores que trabajan para una compañía de productos químicos sencillamente no toman en cuenta los costos públicos de la limpieza de desechos tóxicos, y los investigadores que trabajan para un grupo ambientalista no consideran los efectos de mayores costos de limpieza para la supervivencia de la compañía. El marco mental puede ser automático e inconsciente, o deliberado. Agencias o intereses particulares pueden enmarcar la investigación en formas que probablemente promuevan los valores que ellos defienden. Pero algún conjunto de premisas de valor rodea toda la investigación, y el que consideremos que un estudio se basa en premisas justas o sesgadas es algo que depende más de que estemos de acuerdo con los valores integrados, que de las características del propio estudio.

La investigación que se presenta como argumento en apoyo de una conclusión particular tiene grandes ventajas para los ajetreados tomadores de decisiones: (1) Les ahorra tiempo y trabajo. No tienen que descifrar las implicaciones de la investigación; la investigación como argumento explica las implicaciones de sus descubrimientos en términos fáciles de comprender. (2) El argumento relaciona explícitamente la investigación con el problema en cuestión. Dice: por causa de la información a, b y c, cámbiese el artículo 4 (a) en tal y tal sentido. Esa concreción es muy atractiva. El argumento empieza con la preocupación inmediata del elaborador de la política y presenta la investigación como garantía. (3) La integración de

argumento y evidencia forma un limpio paquete para emplearlo en las negociaciones burocráticas o legislativas. Y los partidarios ya existentes pueden utilizar el paquete en el debate y la negociación.

¿Quién transmitirá, probablemente, la investigación como argumento? Los grupos de interés son un canal. Los grupos de interés y las organizaciones de cabildeo diseminan cantidades considerables de información analítica, con objeto de fortalecer sus posiciones. En un estudio sobre los «cabilderos» de Washington, Schlozman y Tierney (1986) descubrieron que la abrumadora mayoría de los cabilderos (92 %) informaron que la función de «presentar resultados de investigación o información técnica» era parte de su tarea. Entre las actividades a que dedicaron la mayor parte de su tiempo y de sus recursos, los «cabilderos» la colocaron en tercer lugar de las 27 actividades enumeradas. Solo «establecer contacto directo con funcionarios del gobierno» y «atestiguar en las audiencias» requerían más tiempo y esfuerzo que presentar los resultados de una investigación. En mi investigación del Congreso de los Estados Unidos, descubrí que el personal del comité se basaba notablemente en representantes de grupos de interés para obtener información acerca de sus posiciones ante cláusulas legislativas, y también en busca de testimonios para que esas posiciones fuesen convincentes (Weiss, 1989).

Los grupos de interés —grupos de intereses públicos, asociaciones profesionales, grupos de intereses privados— son especialmente notorios en la legislatura. Es allí donde la investigación como argumentos parece florecer más abiertamente. En estos días, cuando el gobierno de los Estados Unidos está dedicado a microadministrar detalles que solían dejarse para que las agencias ejecutivas los aplicaran, los legisladores necesitan prescripciones e información detalladas. También es en la legislatura donde más apremia el tiempo; por eso, si se quiere que la investigación afecte la acción legislativa, parece hacerlo con la mayor facilidad en forma de argumentos.

A primera vista, las burocracias parecen menos receptivas a la investigación como argumentos. Tienen un personal más numeroso y especializado y poseen sus propias oficinas de investigación, análisis y evaluación. Patrocinan su propia investigación interna y externa y se mantienen al día con respecto a los avances ocurridos en su campo. Su periodo también es

más largo, y por ello tienen mejores memorias institucionales de asuntos, investigación y argumentos anteriores. Y para obtener conocimientos no dependen tanto de cómo se los presenten los grupos de interés.

Y sin embargo, a pesar de ser expertos, los burócratas no son inmunes al señuelo de la investigación como argumento. El rasgo distintivo de las agencias burocráticas es lo improbable de que los partidarios de los argumentos no formen parte de los grupos de interés. Más a menudo son partidarios comprometidos que se encuentran dentro de las filas de la burocracia misma. Las facciones internas se valen de la investigación para fortalecer sus argumentos en los debates de la organización.

Las tres imágenes de la investigación de políticas —datos, ideas y argumentos— hacen distintas suposiciones acerca de la naturaleza del proceso de hechuras de las políticas (véase Wittrock, capítulo XV). Bajo la imagen de la investigación como datos se encuentra una visión tecnocrática. Los funcionarios están muy deseosos de hacer un trabajo competente. Hay poco conflicto de metas; los datos de los investigadores son compatibles con las necesidades de los usuarios; los usuarios tienen conocimiento suficiente de los datos para buscar en ellos calidad y aplicarlos eficazmente al problema del momento. Se ve la traza de una burocracia casi weberiana en que la información se utiliza para aumentar la eficiencia.

La idea de la investigación como ideas tiene un sabor menos mecanicista, más humanista. Los funcionarios no necesariamente se orientan a una tarea inmediata. Múltiples problemas y prescripciones complejas compiten por la atención general. Hay espacio para un considerable «juego» en el sistema. La formulación de los problemas que hacen los investigadores no necesariamente coincide con las necesidades inmediatas de quienes toman las decisiones, y estos pueden recordar selectivamente las ideas que más les gustan. Es un lugar de autoridad dispersa. La imagen del modelo de la toma de decisiones como un «cesto de basura» de March y Olsen (1976) es perfectamente compatible con el concepto de la investigación como ideas. Problemas van y vienen; ideas van y vienen; los participantes pueden o no asignar una idea a un problema de momento. Pero las ideas quedan confusamente almacenadas en la memoria de la gente y pueden resurgir al aparecer nuevas cuestiones.

La investigación como argumentación presupone una toma de decisiones con factores en pugna. Tiene un trasfondo político. Intereses y valores se encuentran en disputa. Los partidarios de ciertas ideas utilizan la investigación cuando y si favorece sus propios casos, en el barullo de la toma de decisiones organizativa. Es probable que la investigación beneficie a los intereses en pugna que tengan los recursos necesarios para pagar investigación, y que tengan el refinamiento necesario para utilizar la investigación con el fin de promover su bando en la controversia.

Conclusiones tentativas e hipótesis nuevas

La investigación política se utiliza en forma de datos para toda una gama de usos especializados, pero a menudo son cuestiones detalladas y cotidianas. La influencia de los datos sobre las grandes cuestiones parece relativamente insólita fuera del ámbito de la economía. La investigación como ideas fluye a la arena de las políticas pasando por toda una variedad de fuentes no coordinadas. Este uso parece ser omnipresente, pero no está claro el efecto que tenga sobre elecciones específicas de política. Ideas llegadas de la investigación se vuelven corrientes, parte del discurso político y, sin embargo, no afectan la acción de las autoridades. Tampoco sabemos bien qué importancia tenga la investigación como fuente de ideas, en contrapeso con otras fuentes. La investigación entra en la arena de las políticas como *argumentos*, principalmente por medio de las actividades organizadas de sus partidarios. Los grupos de interés que operan en la legislatura y en las facciones internas que compiten dentro de las agencias ejecutivas pueden hacer uso de la investigación por medio de la argumentación. Citar el apoyo de la investigación a menudo parece ser táctica eficaz para llamar la atención hacia una posición, sobre todo cuando sus partidarios tratan de llevar sus argumentos al público. La evidencia de la investigación da legitimidad al caso y, a veces, mayor notoriedad. Sin embargo, el resultado de cualquier argumento particular puede no depender tanto de la fuerza de la evidencia cuanto del poder de la coalición que la apoya.

Es difícil tratar de desenredar el extraño papel de la investigación, por causa de la constante interacción de información, ideología e intereses. Y la investigación no triunfa. No pasa por encima de creencias e intereses.

Lo que sí puede hacer es ayudar a las personas a figurarse cuál decisión específica será la que más favorezca sus intereses y sus convicciones ideológicas en un caso particular (Weiss, 1983). Dado que es tan complejo el proceso de llegar a una decisión, primero en la mente de cada persona y luego entre individuos, para la organización en conjunto el papel de la información (y de ese subconjunto de información que procede de la investigación) probablemente no podrá saberse nunca con certidumbre.

En cambio, podemos ofrecer una serie de hipótesis. Con base en la investigación efectuada y el testimonio de los participantes interesados, deseo ofrecer estas hipótesis preliminares acerca de la influencia de la investigación en forma de datos, ideas y argumentos sobre la política pública.

La investigación como datos más probablemente tendrá influencia:

"en situaciones de consenso sobre valores y metas. La investigación puede precisar el problema y aclarar sus parámetros, y servir de base para hacer buenas estimaciones de la eficacia de los correctivos;

cuando se plantean claramente dos o tres opciones. Si se ha diseñado explícitamente la investigación para poner a prueba cierto número limitado de opciones y si los descubrimientos son bien claros, los datos tendrán buena oportunidad de establecer una diferencia;

en situaciones en rápido cambio. Cuando nadie sabe cuál es la situación, es probable que se atienda a datos, particularmente si muestran que las condiciones actuales están fuera de la zona de aceptabilidad (Steinbruner, 1974). Sin embargo, más probablemente se les utilizará como «reconocimiento» y como trasfondo que como guía para la acción;

cuando los que toman las decisiones (o su personal) tienen buenos conocimientos analíticos. Utilizar datos es algo que exige un conocimiento de las técnicas de investigación para que se pueda evaluar la calidad de los datos, y una comprensión de las limitaciones de la investigación, de modo que los datos no se apliquen a situaciones que vayan más allá de su generalizabilidad".

La investigación como ideas más probablemente tendrá influencia:

"en las primeras etapas de una discusión política. Es entonces cuando hay espacio para la consideración de las distintas facetas de un problema y para las diversas soluciones. Cuando la discusión ya ha llegado demasiado lejos, quienes toman las decisiones han enmarcado el asunto y empezado a endurecer sus posiciones;

cuando la política existente está en desorden. En condiciones de fracaso y crisis, aun los más endurecidos buscan un modo de salir de la situación;

cuando es grande la incertidumbre. Cuando nadie sabe qué hacer o qué funcionará, las ideas tienen buena demanda;

en las arenas de política descentralizadas, cuando deciden muchos cuerpos separados. En arenas como la justicia penal o la educación, cuando está dispersa la autoridad de las decisiones, una idea relativamente sencilla puede llegar más lejos y con mayor rapidez que unos datos detallados".

La investigación como argumentos más probablemente tendrá influencia:

"cuando es intenso el conflicto. En un área conflictiva donde los diferentes bandos han marcado sus posiciones, lo que cada quien está buscando es garantía y apoyo para fortalecer su propio caso. La investigación puede aportar pruebas que tranquilizarán a los partidarios, convencer a los indecisos y debilitar las posiciones rivales;

en las legislaturas. Como la legislatura es el sitio quintaesenciado para la resolución de diferencias ideológicas y basadas en un interés, la argumentación es el humor prevaleciente, y será bien recibida toda investigación que apoye ciertos argumentos;

después de tomadas las decisiones (como legitimación). La toma de decisiones es la estación de paso a la puesta en marcha y al impacto último. Para unir la organización y los individuos que aplicarán las decisiones, hay una continua necesidad de legitimación. Y la investigación puede ayudar a satisfacer la necesidad".

El dilema normativo

El análisis nos sugiere, hasta aquí, que pueden obtenerse beneficios cuando se emplea la investigación de políticas en un papel de defensa activa en la toma democrática de decisiones. Los primeros beneficios son un aumento de notoriedad e influencia. La investigación utilizada como apoyo para un argumento de políticas no tiene que salir a buscar quién la atienda; encuentra un público ya dispuesto entre quienes están comprometidos con las políticas que la investigación propone. Se alegrarán de divulgar y propagar sus descubrimientos, como armas para su causa. Por ello, es probable que se logre evitar el frecuente destino de la investigación de políticas: languidecer olvidada y al margen.

Además, en cuanto esa investigación entre en la liza de las políticas, probablemente tendrá una mejor oportunidad de surtir efectos sobre la discusión de políticas. Por una parte, refuerza las convicciones de partidarios ya dispuestos y les da mayor confianza en su posición. Si en los descubrimientos de la investigación hay elementos que arrojan dudas sobre ciertas premisas de su posición, podrán fortalecer sus argumentos. También se les podrá atraer a modificar ciertos subpuntos de sus propuestas para hacerlos más congruentes con la evidencia. Armados con testimonios convincentes, también podrán estar mejor equipados para convencer a los indecisos a que se pongan de su lado. Por ello, es probable que puedan fortalecer su coalición y mejorar las posibilidades de que prevalezca la posición favorecida por la investigación.

Aunque la investigación como argumento puede ser más persuasiva en la arena de las políticas, representa una mezcla de funciones científicas y partidaristas. Así, vuelve a plantear ciertas preguntas difíciles acerca del papel apropiado de los expertos en una sociedad democrática y, particularmente, preguntas acerca de la ciencia social como conocimiento científico o argumento especial. Si los científicos sociales abandonan su insistencia en la objetividad, ¿no estarán socavando la posición de neutralidad experta que, para empezar, les valió ser invitados a la arena de las políticas? Esta es una interrogante que tiene ya un largo historial en la ciencia social norteamericana.

Antecedentes históricos

En los Estados Unidos, las ciencias sociales empezaron a separarse de la filosofía moral a mediados del siglo XIX, como respuesta, en gran parte, a los problemas planteados por la Revolución Industrial. Los primeros científicos sociales eran reformadores: sociólogos, politólogos y, sobre todo, economistas. Fueron atraídos a las nascentes disciplinas, primero por causa de su interés en aportar un conocimiento que aliviara las condiciones de pauperismo, inquietud laboral, poblaciones no asimiladas, barrios bajos y otras patologías de los Estados Unidos durante su industrialización en el siglo XIX. Los primeros científicos sociales fueron ciudadanos voluntarios; poco después, estos «aficionados» fueron desplazados por académicos profesionales. Según Furner (1975: 3):

Ambos grupos se enfrentaron a un problema común: cómo reconciliar los intereses reformistas que para empezar les habían atraído a la ciencia social, con su necesidad de reconocimiento como científicos objetivos y desapasionados, capaces de dar una guía desinteresada a la sociedad.

La solución que con el tiempo crearon consistió en subrayar en sus disciplinas los elementos teóricos y de formación de conocimiento. Al profesionalizarse las ciencias sociales y expulsar a sus miembros «aficionados» y al convertir, cada vez más, a la universidad en su base, el hincapié en el conocimiento fue una actitud natural, lo que no significó que los científicos sociales abandonaran el impulso reformista. En los últimos decenios del siglo se elaboró, por medios tentativos, una definición de los límites ideológicos apropiados, y surgió un consenso sobre los límites permisibles de desviación. Los científicos sociales que se quedaron dentro de la corriente liberal aceptable pudieron afirmar entonces que su labor era objetiva y que merecía la protección de la libertad académica. Los científicos sociales «envolvieron sus intenciones reformistas en un manto de prerrogativa profesional, que los protegió de las consecuencias de hacer recomendaciones que de otra manera habrían constituido un gran riesgo» (Furner, 1975: 322). Y la objetividad se convirtió en escudo contra todo ataque. En la academia, los científicos sociales abrazaron la objetividad y el desarrollo de teorías como elementos centrales de sus ciencias de la sociedad. Hasta los científicos sociales que parecían utilizar la investigación para hacer casos de políticas, pudieron contar con su

adherencia a las normas profesionales, para protegerse de contraataques de quienes se encontraban fuera del privilegiado bastión de la disciplina y de la universidad (para una perspectiva comparativa sobre estos desarrollos, véase Wittrock *et al.*, capítulo II).

Los entendimientos negociados a finales del siglo XIX y principios del XX sirvieron bien a los científicos sociales (Lyons, 1969). Pudieron hacer avanzar la base del conocimiento de las disciplinas y al mismo tiempo utilizar su labor en provecho de su agenda sociopolítica. Pero a finales del siglo XX hubo un difundido reconocimiento de la intrusión inevitable de los valores, aun en la investigación más objetiva y desapasionada. Los científicos sociales llegaron a reconocer el hecho de que la actitud política y filosófica de su labor era influida por su restringida visión teórica, sus suposiciones iniciales, sus preferencias metodológicas, sus incompletos modelos explicativos y otros obstáculos inevitables hacia la verdadera objetividad. En realidad, muchos llegaron a cuestionar si había fuera un «mundo real» independiente de la interpretación que ellos daban a sus parámetros. La «construcción social de la realidad» (Berger y Luckmann, 1980) llegó a ser un concepto aceptado.

La conciencia de los múltiples mundos que se ofrecían al estudio de la ciencia social recibió un apoyo inesperado: el de las teorías de la física del siglo XX. El Principio de Incertidumbre de Heisenberg, 1927, afirmaba que los científicos pueden fijar la posición o la velocidad de una partícula subatómica, pero no ambas cosas, porque el propio acto de medición interfiere con lo que se está midiendo. Durante los treinta, Niels Bohr fue el principal creador de la Interpretación de Copenhague de la mecánica cuántica, la cual niega la existencia de todo mundo independiente de lo que se pueda medir. La física, dijo, no puede descubrir lo que es la naturaleza; la física trata de lo que puede decirse acerca de la naturaleza. El esfuerzo de Einstein por salvar la idea de sentido común de un mundo tangible, por la premisa de la localidad, ha resultado vano. Por muy contraintuitiva e ilógica que parezca la mecánica cuántica, hasta la fecha en todos los experimentos se ha descubierto que predice correctamente los fenómenos. Y sin embargo, parece decir que a nivel subatómico no existe algo que pueda llamarse «cosas». Los objetos no tienen propiedades definidas hasta que se les mide. Algunos físicos han propuesto una teoría de «muchos mundos» para explicar el extraño comportamiento de las partículas, que

viajan por universos paralelos (*The Economist*, 1989). El simple y seguro mundo del positivismo se ha ido para siempre.

El hecho de que los científicos sociales moldean el mundo que estudian por la manera en que definen el problema ha llegado a ser aceptado no solo por científicos sociales, sino también por maduros actores políticos. Están conscientes de que las suposiciones, teorías y elección de variables de los investigadores pueden tener grandes efectos sobre las respuestas que encuentren. Esta nueva interpretación pone en duda el acomodo que habían negociado anteriores generaciones de científicos sociales. Si ya no pueden afirmar que descubren la «verdad» acerca de la «realidad», ¿cuál es su función en el proceso de las políticas? Parece haber llegado la hora de formular un nuevo conjunto de suposiciones y disposiciones.

En forma contemporánea y coincidente, los dos decenios pasados presenciaron el surgimiento de las ciencias de las políticas en la universidad norteamericana. Se establecieron nuevas escuelas de política pública; muchas de ellas obtuvieron considerable notabilidad y renombre. En algunas universidades, antiguos departamentos de administración pública se han transformado en escuelas de asuntos públicos, o se han dividido en dos; la política pública/asuntos públicos ha sentado sus reales. Se han fundado nuevas publicaciones y se han organizado nuevas asociaciones profesionales, cuyo número de miembros ha ido en aumento. Los estudios de políticas se han convertido en la subespecialidad en más rápido crecimiento en la ciencia política norteamericana, y el análisis de políticas interdisciplinario (a menudo con la economía en ascenso) se han vuelto sumamente conspicuos no solo en las escuelas de política pública/asuntos públicos, sino también en escuelas profesionales como las de salud pública, trabajo social y educación (véase Jann, capítulo IV). Y el movimiento no se ha limitado a las universidades. Después de un breve recorte a comienzos de los ochenta, los gobiernos han seguido contratando analistas de políticas para trabajar en agencias ejecutivas y legislaturas, a los niveles federal y estatales, y hasta en algunas de las ciudades más grandes. Los analistas de políticas han servido como asesores de agencias y de miembros de comisiones de alto nivel. Los gobiernos siguen invirtiendo sumas considerables de dinero en investigación de políticas dirigida por organizaciones de investigación externas.

¿El papel de abogado?

Toda esta actividad parece descansar en suposiciones no reflexionadas acerca de la objetividad de la ciencia de las políticas. Pero si las viejas pretensiones de objetividad se han desacreditado, ¿habrá llegado el momento de crear una nueva interpretación del papel de la investigación de políticas? ¿O volverá a desencadenar semejante revisión de las reglas básicas las batallas que entablaron antiguas generaciones de científicos sociales y (evidentemente) ganaron?

Resulta tentador buscar una resolución fácil al dilema: mantener a los investigadores y su investigación en categorías separadas. La investigación debe permanecer tan objetiva como sea humanamente posible. Entonces, los investigadores podrán utilizar la investigación para toda una variedad de propósitos, uno de los cuales —si así lo prefieren— puede ser el de la defensa activa, la promoción y las recomendaciones. Pero esa fácil distinción simplemente confunde la complejidad del asunto. Si la interpretación pospositivista es que no hay una sola realidad (o absolutamente ninguna realidad, hasta que tomemos su medida), una objetividad «humanamente posible» resulta bastante mediocre.

Para avanzar en nuestro análisis, tomemos en cuenta dos factores. El primero es aquello que los investigadores de las políticas desean realizar en sus funciones profesionales. ¿Hasta qué punto están interesados en la defensa activa y la formulación de recomendaciones, o las rechazan de antemano? El segundo tiene que ver con la ubicación diferencial de los investigadores de políticas. ¿Hasta qué punto comparten estos en las universidades las oportunidades y los riesgos con quienes trabajan dentro del gobierno o en organizaciones de investigación independientes? Tal vez haya que hacer distinciones, dependiendo de la afiliación institucional de quienes emprenden la investigación de políticas. Así como el acomodo del siglo XIX dependió marcadamente de la creación de departamentos de ciencias sociales en las universidades, que podían ofrecer a la vez una razón pedagógica y un refugio para los científicos sociales reformistas, así un actual acomodo puede depender de las estructuras dentro de las cuales trabajan los investigadores orientados hacia las políticas.

Las intenciones de los investigadores de las políticas

¿Qué esperan lograr los investigadores de las políticas? No conozco datos al respecto, pero en el estudio informal que he estado haciendo descubrí cuatro motivaciones básicas; creo que representan las metas de los investigadores que trabajan dentro del gobierno, de quienes trabajan en organizaciones de investigación, lucrativas y no lucrativas, y de quienes trabajan en institutos y departamentos universitarios.

Estas son las cuatro metas que al parecer tienen a la vista los investigadores de las políticas:

- 1) Un objetivo es llegar a ser un conocido y respetado científico social gracias a la calidad conceptual y metodológica de su investigación. Los investigadores de las políticas desean realizar bien su trabajo, quieren estar al día, enterados de los últimos métodos y de las teorías más pertinentes, y aspiran a ganarse el respeto de sus colegas y superiores así como de las personas para quienes se realiza la investigación.
- 2) Una segunda meta consiste en establecer una diferencia. Desean que su investigación influya sobre las políticas, que sea «utilizada». El principal atractivo de la investigación de las políticas es estar cerca de la acción y participar en la formulación de las políticas. Y a menos que su trabajo tenga público, toda la empresa parece carente de significado. Los investigadores de las políticas desean servir para algo.
- 3) Un tercer objetivo es llegar a ser lo que Schultze (1968) llamó *abogados partidaristas de la eficiencia en el gobierno*, es decir, favorecer la causa de una toma de decisiones analíticamente basada. Quieren ver que esa investigación sirva para hacer más eficiente al gobierno al alcanzar unos fines políticamente determinados. Los abogados de la eficiencia no representan una solución particular, sino la aplicación de datos, cálculos y racionalidad en el gobierno. Su meta es hacer que todo el sistema sea más racional, aclarando el planteamiento de problemas, comprendiendo la envergadura y distribución de los problemas, y evaluando sistemáticamente los costos y beneficios de las diversas opciones.

- 4) Un cuarto objetivo consiste en defender una posición política o una perspectiva política y plantear esa perspectiva en la arena de las decisiones. Quienes tienen este objetivo tratan de promover una política específica o una orientación general hacia las políticas, sobre la base de sus propios valores. Si trabajan en la educación, acaso deseen impulsar las políticas hacia una mayor gama de oportunidades para padres y estudiantes, sobre a cuáles escuelas asistirán los hijos. Si trabajan en la energía, acaso deseen promover la energía solar. Hacen investigaciones que favorecen su opción predilecta. Aquellos cuya posición se deriva de sus creencias, sean las suyas o las de los clientes a los que sirven y que utilizan la investigación para favorecer esas creencias, pertenecen a la categoría de los abogados defensores.

En mi somera revisión, la respuesta más frecuente a la pregunta de lo que desean realizar los investigadores de las políticas es desean que se utilice su investigación. Anhelan sentirse eficaces. Aun quienes dicen que desean ser buenos investigadores, debidamente respetados, a menudo consideran importante esa meta porque tiende a aumentar la aceptación de su investigación entre quienes toman las decisiones y, por tanto, su utilización. Hasta la fecha, el objetivo menos mencionado ha sido promover los propios valores y posiciones de políticas. Solo unos cuantos investigadores de afiliación universitaria han mencionado esto como parte de su intención. Aunque la escasez de esta respuesta puede deberse a simple disimulo, creo que probablemente la defensa y promoción es relativamente rara como objetivo consciente. Los investigadores de las políticas están dispuestos a reconocer que, indudablemente, se hallan influidos por sus valores al enmarcar la investigación, formular preguntas, interpretar datos y la mirada de otras decisiones de investigación que tienen que tomar. Pero se esfuerzan en serio por ser tan objetivos como les sea posible, en parte porque tal es la norma de trabajo, pero —de mayor importancia— porque no desean enajenarse grandes partes de su público adhiriéndose a una posición ideológica. Fueron invitados a participar en la arena de las políticas porque pueden aportar evidencias, no porque califiquen como «jugadores», y toman en serio su tarea. También insisten en que estudian muchas cuestiones en que no tienen opiniones o puntos de vista previos, y todo lo que saben es lo que aprenden de su investigación.

Variaciones por ubicación

Es indiscutible que la constelación de motivos varía según la ubicación. Los investigadores de políticas que trabajan directamente para una organización que defiende, promueve y hace recomendaciones (como un sindicato o un grupo de interés) se han comprometido a emplear la investigación como parte de su argumentación; desde luego, esto no significa que tengan que deformar su investigación para ponerla en «armonía» con su organización, o que no exijan —a veces a gritos— un cambio, cuando los resultados de la investigación no apoyan la posición que está adoptando la organización. Pero semejante trabajo incluye, como parte del contrato implícito, una aceptación de la premisa de la investigación en apoyo de las tareas de defensa activa y formulación de recomendaciones.

Por otra parte, los investigadores de las políticas que trabajan en el gobierno suelen tener un interés en subrayar la objetividad de su labor. En particular, si trabajan en una dependencia en que están divididas las opiniones, o para una legislatura en que está representada toda la gama de la opinión política, comprenden que su «boletín de entrada» depende de que se les considere imparciales y de amplio criterio. La objetividad perfecta tal vez sea imposible, pero los analistas pueden dar pasos extraordinarios para apartar de su trabajo todo sesgo personal. Por ejemplo, en la Oficina Congressional del Presupuesto, los revisores leen cada manuscrito palabra por palabra para ver que cada afirmación esté basada y apoyada por un testimonio específico.

Una actitud de neutralidad hacia políticas alternativas sirve a varios propósitos vitales. El fundamental es lograr que la dependencia acepte la investigación considerando que aporta motivos neutrales y convincentes para la toma de posiciones. A la premisa subyacente se le puede llamar, según el término de Miles y de Huberman, *positivismo lógico blando* (1984: 19). Es el positivismo lógico que «reconoce y trata de compensar las limitaciones de este enfoque» (p. 19). Pero desde luego, dentro de una dependencia administrativa, algunas interpretaciones pueden ser tan omnipresentes que influyan sobre la investigación, así como sobre todas las demás actividades. Por ejemplo, en una agencia de salud mental, los enfoques de la comunidad hacia la atención a la salud mental pueden ser

considerados tan generalmente valiosos y vitales, que toda la investigación de la agencia les dé prioridad.

Los investigadores de políticas que trabajan en organizaciones de investigación, sean lucrativas o no lucrativas, probablemente compartan muchas de las creencias de quienes trabajan en el gobierno. Si los fondos para su investigación dependen de organismos gubernamentales, tendrán los mismos incentivos para mantener una actitud de neutralidad hacia las políticas, de modo que su investigación parezca persuasiva no solo a sus clientes, sino también a los grupos exteriores a quienes sus clientes deseen convencer. Al mismo tiempo, probablemente habrá una ventaja en aceptar las premisas básicas de la agencia patrocinadora (por ejemplo, que la promoción de la salud mental comunitaria reciba la mayor prioridad). Si existe un incentivo especial para las organizaciones investigadoras, consiste en no enajenarse a otros potenciales patrocinadores. Una vez más, se haría cierta presión para transmitir cierta aura de objetividad.

En la universidad, los investigadores de política tienen mayor espacio. En parte por causa de las victorias por la libertad académica obtenidas en los primeros años del siglo, en parte por la justificación pedagógica para su posición, en parte porque dependen más del reconocimiento de sus compañeros y, a la vez, dependen menos de que los funcionarios *utilicen* su investigación para validarla, los investigadores universitarios pueden adoptar más fácilmente una orientación de políticas basada en sus propios valores y utilizar su investigación como parte de la argumentación. Hasta cierto punto, esto significa que pueden permitirse ser sinceros y francos acerca de sus valores, en lugar de ocultarlos hasta el grado que sus colegas no universitarios consideran esencial. De hecho, todos los pocos investigadores de políticas que respondieron a la encuesta informal diciendo que trataban de favorecer particulares puntos de vista ocupaban puestos universitarios. Los investigadores pueden, en las universidades, defender posiciones controvertibles y actitudes de crítica a las políticas y las premisas actuales, casi sin temor a ser castigados. Existen tradiciones de defensa activa y recomendación, precedentes de crítica y recompensas para las líneas de pensamiento originales. Dado que adoptar actitudes y presentar evidencia en apoyo de dichas actitudes es algo aceptado y hasta protegido, los investigadores universitarios pueden promover teorías y políticas particulares.

El hecho de que *puedan* hacerlo no significa, desde luego, que muchos investigadores académicos *lo hagan*. En realidad, hasta en la academia existe una considerable resistencia a la idea de la investigación como argumentación. No pocos investigadores de políticas han declarado, por escrito, que usar la investigación para favorecer una posición preexistente es un «abuso» de la investigación (Horowitz, 1975; Heller, 1986). Se supone que primero viene la investigación y luego las implicaciones deberán tomarse de los datos; adaptar los descubrimientos a una postura de políticas preconcebida es algo sospechoso. Desde luego, parecerá mucho más sospechoso si la posición preexistente está siendo defendida por alguien más, ya sea la Asociación Nacional del Fusil o Control de Armas de Mano, SA. Pero aunque sean las preferencias en materia de políticas del propio investigador las que se defiendan, esto puede parecer tendencioso... a menos que la preferencia de políticas sea inobjetable porque caiga dentro de la corriente liberal que tiende a dominar la ciencia social académica.

Pros y contras de la investigación como argumento

Y sin embargo, hay algo renovador en utilizar la investigación como parte de un argumento de políticas. Siempre que la investigación no sea deformada, parece apropiado para sus partidarios utilizar su testimonio en favor de sus argumentos. La investigación puesta al servicio de la argumentación hace descender la investigación de su pedestal por encima de la mayoría y la reconoce como una fuente de conocimiento entre muchas otras. Podría afirmarse que esta es una presentación más honrada: que la investigación, que inevitablemente es parcial en su planteamiento, conducta e interpretación, llega a su estado apropiado como apoyo parcial de una postura partidarista parcial. Escribe Rivlin (1973: 62):

La ciencia social forense [en que los especialistas se arrojan la tarea de escribir informes en favor de posiciones particulares de políticas] me parece un desarrollo sumamente saludable. Reduce la hipocresía de la seudobjetividad y los sesgos disimulados. Si se le emplea bien, puede agudizar las cuestiones públicas y hacer que la investigación científica social sea más pertinente para cuestiones reales de lo que ha sido en el pasado.

Lindblom y Cohen (1979: 81) han escrito: «La complejidad de los problemas niega la posibilidad de toda prueba y reduce la búsqueda de los hechos

a la búsqueda de aquellos hechos selectivos que, si son debidamente desarrollados, constituyen un testimonio en favor del argumento relevante». Se refieren a la discusión que presenta Majone (1977) del análisis de políticas, en que afirma que este desempeña funciones posdecisión, como mostrar lo racional de una decisión, recomendar la decisión a otros y ponerla en un contexto de pensamiento racional (Lindblom y Cohen, 1979: 80). Majone, como Lindblom y Cohen, considera apropiados estos usos, aunque establece el punto de que la investigación tiene que poner a prueba, además de hacer embonar el argumento.

Lindblom ha sostenido desde hace largo tiempo que las más de las veces, y muy sensatamente, la investigación se utiliza de manera partidarista y en apoyo de fines partidaristas (Lindblom, 1980). Ha escrito (p. 32):

Aunque gente partidarista solo utilizará los recursos analíticos para favorecer sus propios intereses, interactúa con muchos otros partidarios que hacen lo mismo en favor de sus intereses respectivos. Como cada quien alimenta los procesos interactivos con análisis para su propio beneficio, al menos parte de él se vuelve posesión común de todos los participantes (...). Muy posiblemente, la supuesta «competencia de ideas» considerada esencial para la democracia pluralista, en gran parte tome la forma de un intercambio entre analistas partidarios y elaboradores de políticas partidarias, que utilizan su análisis.

De este modo, en el toma y daca entre partidarios, el análisis partidarista puede llegar a ser una ayuda para el desarrollo de una política consensual. Semejante opinión representa, sin duda, una interpretación optimista de las posibilidades, como al parecer lo reconoce Lindblom, pero no parece demasiado tirado de los cabellos.

Yo simpatizo con la investigación como argumentación, en términos muy parecidos a los de Rivlin, de honradez y pertinencia. Hasta parece justificable cuando la posición de políticas que se está defendiendo no solo representa las preferencias del propio investigador, sino también es apoyada por otros factores en la arena de las políticas, aun cuando las propuestas así apoyadas son anteriores a la investigación y no se derivan básicamente de sus descubrimientos. Me parece que las principales condiciones son que no se deben deformar los descubrimientos y que todo el *corpus* de evidencia debe estar a disposición de todas las partes

que participen en el debate. Dentro de esos límites, me parece que el uso de la investigación como testimonio para promover una medida política particular mantiene limpio el juego, al menos en la superficie, mientras que al mismo tiempo extiende la gama de conocimiento que se toma en cuenta en el desarrollo y la adopción de las políticas.

Pero hay costos. Un costo inmediato es que algunos bandos en los debates de políticas tienen mayor habilidad y más recursos para aprovechar la investigación. Quienes no tienen experiencia en la investigación, o cuya experiencia les ha llevado a denigrar o desconfiar de la investigación, difícilmente se aprovecharían de los recursos de investigación. Los que no tienen fondos ni aliados entre los investigadores de las políticas u otros medios de adquirir investigación se encontrarán en desventaja. Mientras no se rectifiquen esos desequilibrios, la investigación no será un árbitro neutral, sino una ayuda para el bando privilegiado.

Un costo que se paga a más largo plazo consiste en que reinaugura la vieja pugna acerca del papel apropiado para las ciencias sociales en la vida social y política. Si los científicos de las políticas aceptan la situación de participar en la defensa y de promover recomendaciones y abandonan las pretensiones de verdad y de objetividad, estarán participando en un desarme unilateral. Dejan de depender de su arma más poderosa (la evaluación minuciosa y sistemática de la evidencia empírica pertinente) y tienen que jugar de acuerdo con las reglas del otro. Y los otros tienen poder, persuasión, experiencia, dinero y votos: recursos que no abundan en los círculos de la ciencia social. Casi no cabe duda de cuál será el resultado.

Pero, ¿cómo, con la conciencia tranquila, pueden los investigadores de las políticas sostener que ellos representan la verdad y la objetividad? Están demasiado conscientes de los defectos de esas afirmaciones para exigir que se tenga perfecta confianza en los descubrimientos de su investigación o en las recomendaciones que basan en sus investigaciones. Y también sus clientes conocen esas limitaciones. ¿Qué pueden hacer?

Me parece que una distinción crucial es la diferencia entre mayor y menor objetividad. Hay una diferencia entre la investigación que aspira a ser un análisis justo y completo de un aspecto y la investigación que se pone

en marcha —sea conscientemente o por falta de reflexión y autocrítica— para «defender el caso» de una posición. Un hito de la ciencia es el intento por refutar o, según el término de Popper, falsear las suposiciones de alguien. El esfuerzo por poner a prueba, por desafiar, por tratar de refutar premisas iniciales, es el requisito expreso del enfoque científico y, por tanto, de la investigación de políticas. Si la investigación produce unos descubrimientos que convencen al investigador de la superioridad de un curso particular, eso estará muy bien. Pero empezar con la intención de probar un caso es función del ideólogo, del cabildero o del pistolero a sueldo, no del investigador de políticas.

Pero —y este es un gran pero— no es posible desconfirmar los valores. El investigador que analiza datos contra criterios de eficiencia económica puede ser comprensivo, imparcial y laborioso, pero aun así seguirá favoreciendo cierto tipo de prescripción para una política. Otro investigador que analice los mismos datos con la misma escrupulosidad, pero que sostenga valores de distribución equitativa del ingreso, llegará a conclusiones distintas. Cada uno podrá, con la conciencia tranquila, utilizar el testimonio como argumentos de políticas. Hemos vuelto al antiguo pantano de las disputas de la filosofía de la ciencia sobre los hechos y los valores.

Lo que puede sacarnos del pantano, creo, es la naturaleza del consenso norteamericano. En términos generales, esta no es una sociedad que tenga posiciones violentamente divergentes y polarizadas¹. Los republicanos más conservadores se preocupan, hasta cierto grado, por la situación de los que carecen de hogar y por la salud y educación de los nuevos inmigrantes de color. Los demócratas más liberales se preocupan por lo competitivo de la industria norteamericana y por la reducción del déficit presupuestario. Los actores de la política a lo largo de toda la gama responden a muchos de los mismos valores, si bien con diferente intensidad, y es posible convencerlos de que consideren las propuestas de políticas a la luz de toda una gama de criterios (las excepciones son temas como el aborto, hacia el cual se sostienen opiniones extremas, y los extremistas comparten pocos valores). Dado un amplio y bien recorrido tramo del camino, los investigadores de políticas pueden trabajar a partir de su propio conjunto de valores, sin perder pertinencia en la discusión en general. Tal vez pueda ocurrir algo similar a la «competencia de ideas» de Lindblom.

Pensemos en cómo los periodistas se enfrentan a un problema similar. Los periodistas saben tan bien como los científicos sociales que *sus* informes no son espejos objetivos de alguna «realidad» que está ahí. Más que la ciencia social, la noticia es una construcción social, aun con menos reglas y límites metodológicos. Y sin embargo, los periodistas conservan el principio de objetividad y diariamente tratan de cumplir con normas de precisión e imparcialidad. Los científicos de las políticas tienen muchos mayores recursos teóricos y metodológicos, en su busca de una explicación válida. El hincapié en la evidencia y la lógica protege contra las deformaciones muy burdas. Muchos distintos investigadores de políticas, cada uno con sus prejuicios y limitaciones y con un número mayor de canales para transmitir su trabajo que el limitado número de medios de comunicación, pueden construir *colectivamente* una explicación rica y compleja del mundo.

Y sin embargo, algunos segmentos de la opinión parecen creer que la investigación de políticas no satisface sus intereses. Probablemente, los conservadores forman uno de esos grupos. Por ejemplo, Banfield (1980) afirma que el análisis de políticas intrínsecamente encuentra fallas en las instituciones y localiza nuevos problemas que necesitan solución del gobierno. Como lo mostraron los acontecimientos de los ochenta, los gobiernos conservadores, al tomar posesión, suelen recortar los presupuestos de la ciencia social. Esto ha ocurrido en los Estados Unidos y también en la Europa occidental (véase Wittrock *et al.*, capítulo II). Ciertamente es que los científicos sociales tienen una orientación más liberal que la población en general (Orlans, 1973; Lipset y Ladd, 1972). El impulso reformista que animó a las nacientes ciencias sociales en el siglo XIX probablemente resuena hoy en las ciencias de las políticas, y puede tender a enajenarse a quienes tienen ideas más conservadoras².

Puede aceptarse que la inclinación colectiva de la investigación de políticas sea por lo dado —lo dado compasivo— en la situación. A la investigación se le puede contener a la manera en que la Oficina Congressional del Presupuesto, así como ciertas otras agencias, ejercen su control: por medio de continuas revisiones a la precisión de los datos y la lógica de la interpretación. La política de la dependencia exige que se compruebe toda afirmación empírica que haya en un informe. O bien —y esta es

una opción que suele gustarme, para quienes pueden darse el lujo— los investigadores de las políticas pueden hablar con franqueza acerca de sus valores y sus posiciones desde el principio. Si empiezan por mostrar sus valores, los lectores podrán descomponer en factores los sesgos que haya en su interpretación de lo que dice la investigación, aplicando cualquier «factor de corrección» que requieran sus propias inclinaciones³. Cuando se conoce la posición del investigador, los lectores se ven invitados a leer con mirada crítica y a revisar las implicaciones de la investigación y las recomendaciones, con ojo experto.

O tal vez haya llegado el momento de volver a enfrentarnos a los antiguos dilemas del conocimiento y la participación activa para llegar a un nuevo acomodo pospositivista.

Notas

* Título Original “Social Sciences and Modern States”. Primera edición en inglés ©1991 por University of Cambridge Press. Primera traducción al español “*La investigación de políticas: ¿datos, ideas y argumentos?*”, en Peter Warner, Carol Weiss, Björn Wittrock y Helmut Wollman (comp.), “Ciencias sociales y estados modernos: experiencias nacionales e incidencias teóricas”, pp. 327-356, Fondo de Cultura Económica, México, 1999. Reproducción autorizada.

1. La misma pauta puede sostenerse en otras democracias occidentales.
2. Aaron (1978) afirma que el efecto agregado de la investigación de políticas consiste en plantear nuevas preguntas y dudas y, con ello, en hacer más lento el ritmo de la reforma, teniendo un efecto conservador sobre las políticas. Otros afirman que los investigadores de las políticas a menudo adoptan el punto de vista de sus clientes, las agencias del gobierno, actitud que no necesariamente es liberal. No hay duda de que ambas afirmaciones son ciertas en algunas circunstancias. Pero en general se cree que los analistas individuales tienden a adoptar una perspectiva liberal, o liberal de izquierda, en particular quienes ocupan puestos universitarios. Hace algún tiempo, Coleman hizo una tabulación (aunque no he podido encontrarla en letras de molde) mostrando que las recomendaciones que aparecían en los estudios de políticas, más a menudo favorecían los intereses de los clientes de las agencias que los de los administradores de las mismas agencias. Al menos los estudios de políticas que se han publicado en libros y artículos periódicos, o que han aparecido en los principales medios informativos, parecen mostrar una perspectiva a la izquierda del centro.
3. Gans, por su parte, no solo lo ha recomendado, sino también lo ha practicado en su propio trabajo (por ejemplo, Gans, 1979: 303).

Referencias

- Aaron, Henry, 1978. *Politics and the Professors*. Washington, DC, Brookings Institution.
- Alkin, M. C., Richard Daillak y Peter White, 1979. *Using Evaluations: Does Evaluation Make A Difference?* Beverly Hills, Sage.
- Banfield, E. C., 1980. «Policy Science as Metaphysical Madness». En *Bureaucrats Policy Analysts, Statesmen: Who Leads?*, pp. 1-19. Ed. R. A. Godwin. Washington, DC, American Enterprise Institute.
- Banting, K. G., 1979. *Poverty, Politics and Policy*. Londres, Macmillan.
- Barber, Bernard, 1987. *Effective Social Science: Eight Cases in Economics Political Science, and Sociology*. Nueva York, Russell Sage Foundation.
- Berger, Peter L., y Thomas Luckmann, 1980. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*, Nueva York, Irvington (publicado por primera vez en Garden City, Doubleday, 1966). [Hay traducción al español en Amorrortu].
- Bulmer, Martin (ed.), 1978. *Social Policy Research*. Londres, MacMillan.
- , 1982. *The Uses of Social Research: Social Investigation in Public Policy-Making*. Londres, George Allen and Unwin.
- (ed.), 1987. *Social Science Research and Government: Comparative Essays on Britain and the United States*. Cambridge University Press.
- Caplan, Nathan, 1977. «A Minimal Set of Conditions Necessary for the Utilization of Social Science Knowledge in Policy Formation at the National Level». En *Using Social Research in Public Policy Making*, pp. 183-197. Ed. C. H. Weiss. Lexington, MA, Lexington-Heath.
- Cherns, Albert, 1979. *Using the Social Sciences*. Londres, Routledge and Kegan Paul.
- Deitchman, S. J., 1976. *Best-Laid Schemes: A Tale of Social Research and Bureaucracy*. Cambridge, MA, MIT Press.
- Derthick, Martha, y Paul J. Quirk, 1985. *The Politics of Deregulation*. Washington, DC, Brookings Institution.
- Dockrell, W. B., 1982. «The Contribution of National Surveys of Achievement to Policy Formation». En *Social Science Research and Public Policy-Making*, pp. 55-74. Ed. D. B. P. Kallen, G. B. Kosse, H. C. Wagenaar, J. J. J. Kloprogge y M. Vorbeck. Windsor, Berks, NFER-Nelson.
- Elmore, Richard F., 1986. «Graduate Education in Public Management: Working the Seams of Government». *Journal of Policy Analysis and Management*, vol. 6, núm. 1, pp. 69-83.
- Feller, Irwin, 1986. *Universities and State Governments: A Study in Policy Analysis*. Nueva York, Praeger.

- Furner, Mary O., 1975. *Advocacy and Objectivity: A Crisis in the Professionalization of American Social Science, 1865-1905*. Lexington, University of Kentucky Press.
- Gans, Herbert J., 1979. *Deciding What's News*. Nueva York, Pantheon.
- Hayes, C. D. (ed.), 1982. *Making Policies for Children: A Study of the Federal Process*. Washington, DC, National Academy Press.
- Heller, Frank (ed.), 1986. *The Use and Abuse of Social Science*. Londres, Sage.
- Horowitz, I. L. (ed.), 1975. *The Use and Abuse of Social Science*, 2a. ed. New Brunswick, Transaction Books.
- Husén, T., y M. Kogan (eds.), 1984. *Educational Research and Policy: How Do They Relate?* Oxford, Pergamon Press.
- Kallen, D. B. P., G. B. Kosse, H. C. Wagenaar, J. J. J. Kloprogge y M. Vorbeck (eds.), 1982. *Social Science Research and Public Policy-Making*. Windsor, Berks, NFER-Nelson.
- Knorr, Karin D., 1977. «Policymakers' Use of Social Science Knowledge: Symbolic or Instrumental?» En *Using Social Research in Public Policy Making*, pp. 165-182. Ed. C. H. Weiss. Lexington, MA, Lexington Books.
- Lazarsfeld, P. F., W. H. Sewell y H. L. Wilensky (eds.), 1967. *The Uses of Sociology*. Nueva York, Basic Books.
- Leff, Nathaniel H., 1985. «The Use of Policy-Science Tools in Public-Sector Decision-Making: Social Benefit-Cost Analysis in The World Bank». *Kyklos*, vol. 37, pp. 60-76.
- Lindblom, C. E., 1980. *The Policy Making Process*. Englewood Cliffs, NJ, Prentice-Hall (publicado por primera vez en 1968). [Hay traducción al español por el Ministerio para las Administraciones Públicas, de Madrid, y Miguel Ángel Porrúa, de México].
- , y D. K. Cohen, 1979. *Usable Knowledge*. New Haven, Yale University Press.
- Lippitt, Ronald, 1969. «The Process of Utilization of Social Research to Improve Social Practice». En *The Planning of Change*, pp. 142-146. Ed. W. G. Bennis, K. D. Benne y R. Chin. 2a. ed. Nueva York, Holt, Rinehart and Winston.
- Lipset, S. M., y E. C. Ladd, Jr., 1972. «The Politics of American Sociologists». *American Journal of Sociology*, vol. 78, pp. 67-104.
- Majone, Giandomenico, 1977. «The Uses of Policy Analysis». En *The Future and the Past: Essays on Programs*. Nueva York, Russell Sage Foundation. [Traducido al español en Luis F. Aguilar, *La hechura de las políticas*, México, Miguel Ángel Porrúa, 1992].
- March, J. G., y J. P. Olsen, 1976. *Ambiguity and Choice in Organizations*. Bergen, Universitetsforlaget.
- Miles, Matthew B., 1964. «On Temporary Systems». En *Innovation in Education*, pp. 437-90. Ed. M. B. Miles, Nueva York, Teachers College Bureau of Publications.

- , y A. M. Huberman, 1984. *Qualitative Data Analysis*. Beverly Hills, Sage.
- Orlans, Harold, 1973. *Contracting for Knowledge*. San Francisco, Jossey-Bass.
- Patton, M. Q., P. S. Grimes, K. M. Guthrie, N. J. Brennan, B. D. French y D. A. Blyth, 1977. «In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research». En *Using Social Research in Public Policy-Making*, pp. 141-163. Ed. C. H. Weiss. Lexington, MA, Lexington Books.
- Rich, Robert F., 1977. «Use of Social Science Information by Federal Bureaucrats: Knowledge for Action Versus Knowledge for Understanding». En *Using Social Research in Public Policy Making*, pp. 199-233. Ed. C. H. Weiss, Lexington, MA, Lexington Books.
- , 1981. *Social Science Information and Public Policy Making*. San Francisco, Jossey Bass.
- Rivlin, Alice M., 1973. «Forensic Social Science». *Harvard Educational Review*, vol. 43 (febrero), pp. 61-75.
- Sabatier, P. A., 1978. «The Acquisition and Utilization of Technical Information by Administrative Agencies». *Administrative Science Quarterly*, vol. 23, pp. 386-411.
- Saxe, Leonard, 1986. «Policymakers. Use of Social Science Research: Technology Assessment in the U.S. Congress». *Knowledge: Creation, Diffusion, Utilization*, vol. 8, núm. 1, pp. 59-78.
- Scholzman, Kay L., y J. T. Tierney, 1986. *Organizational Interests and American Democracy*. Nueva York, Harper and Row.
- Schultze, C. E., 1968. *The Politics and Economics of Public Spending*. Washington, DC, Brookings Institution.
- Schweinhart, L. J., y D. P. Weikart, 1980. *Young Children Grow Up: The Effects of the Perry Preschool Program on Youths Through Age 15*. Ypsilanti, MI, High/Scope Education Research Foundation.
- Steinbruner, J. D., 1974. *The Cybernetic Theory of Decision: New Dimensions of Political Analysis*. Princeton University Press.
- The Economist*, 1989. «The Queerness of Quanta». 7 de enero, pp. 71-74.
- Walker, Robert, 1987. «Perhaps Minister: The Messy World of "In House" Social Research». En *Social Science Research and Government*, pp. 141-165. Ed. Martin Bulmer. Cambridge University Press.
- Weiss, C. H. (ed.), 1977. *Using Social Research in Public Policy-Making*. Lexington, MA, Lexington Books.
- , 1980. «Knowledge Creep and Decision Accretion». *Knowledge: Creation, Diffusion, Utilization*, vol. 1, núm. 3, pp. 381-404.

Weiss, C. H., 1983. «Ideology, Interests, and Information: The Basis of Policy Positions». En *Ethics, The Social Sciences, and Policy Analysis*. Ed. Daniel Callahan and y Bruce Jennings. Nueva York, Plenum Press.

——, 1986. «The Circuitry of Enlightenment». *Knowledge: Creation, Diffusion, Utilization*, vol. 8, núm. 2, pp. 274-281.

——, 1989. «Congressional Committees as Users of Analysis». *Journal of Policy Analysis and Management*, vol. 8, núm. 3, pp. 411-431.

——, y M. J. Bucuvalas, 1980. *Social Science Research and Decision-Making*. Nueva York, Columbia University Press.

Wilson, J. Q., 1978. «Social Sciences and Public Policy: A Personal Note». En *Knowledge and Policy: The Uncertain Connection*, pp. 82-92. Ed. L. E. Lynn, Washington, DC, National Academy of Sciences.

Raíces de la evaluación: una perspectiva internacional*

Fred Carden y Marvin C. Alkin**

Este artículo considera a los evaluadores o a los grupos de evaluación del mundo en desarrollo y analiza cómo podrían ubicarse en una perspectiva analítica conceptual que represente a los teóricos de la evaluación. El punto de partida para esta reflexión es «el árbol de la teoría de la evaluación» descrito en «Las raíces de la evaluación» (Alkin, 2004, 2012). En una nueva edición de esa publicación, se colocó a los teóricos de América del Norte, Europa y Australasia en un árbol de la teoría de la evaluación conformado por tres ramas principales: métodos, valoración y uso; la rama donde se ubica representa el principal énfasis que el autor le da respecto de las tres ramas. Este capítulo reflexiona sobre la incorporación de las perspectivas de evaluación de otras partes del mundo. El estado de la evaluación en los países con ingresos bajos y medios (PIBM) sugiere dos cuestiones: primero, gran parte de la evaluación se adecua eficazmente al árbol de la teoría como se lo presenta, tal vez con cierto matiz, aunque observamos un origen colectivo más que individual de los contados escritos sobre teoría de la evaluación que descubrimos en los PIBM. Segundo, la naturaleza del trabajo de la evaluación formal en los PIBM es tal que los orígenes prácticos todavía no están formalizados en una teoría prescriptiva completa. Como afirma Lewin (mejor descrito en Marrow, 1969), una buena teoría se construye a partir de la práctica. Si bien la construcción real de la teoría es incipiente en el campo de la evaluación en los PIBM, este estado naciente de la evaluación plantea un primer ejercicio para la definición de qué es una evaluación para el desarrollo.

Teoría de la evaluación/Teóricos

¿A qué nos referimos cuando hablamos de una «teoría» de la evaluación y de quienes la postulan como «teóricos»? En general, hemos llegado a usar el término «teoría» como sinónimo de enfoque o modelo. Entonces no hablamos de una teoría descriptiva o empírica —que describiría lo que ocurriría necesariamente dados conjuntos particulares de acciones—. Existe una escasez de investigación sobre la evaluación. Sencillamente, no conocemos (de un modo científico) los resultados de las acciones de evaluación particulares que tienen lugar en escenarios contextuales particulares. La investigación no proporciona esas respuestas. Tal vez, la complejidad de los contextos de los programas y de las acciones de los evaluadores es tan vasta que jamás seremos capaces de crear una teoría de la evaluación descriptiva.

Como consecuencia de esta deficiencia en la investigación, se desarrolló la necesidad de que hubiera una orientación sobre cómo efectuar una evaluación en el campo. Múltiples autores destacados en la materia aprovecharon la oportunidad para describir su prescripción de cómo se debería realizar una evaluación. Denominamos a estas prescripciones, que se consideran totalmente desarrolladas, «teorías prescriptivas». Entre los nombres vinculados con estas teorías, encontramos a Michael Scriven, Robert Stake, Ernest House, Michael Patton, Jennifer Greene, Yvonna Lincoln, David Fetterman y Barry MacDonald. En cada caso, la identificación de una persona como «teórico» —y de sus escritos como teorías— implica que el trabajo fue lo suficientemente desarrollado y reconocido como para constituir un enfoque distinto.

El árbol de la teoría de la evaluación

La clasificación del árbol de la teoría presentada por Alkin y Christie en el libro «Las raíces de la evaluación» (2004, 2012) brinda una perspectiva analítica rectora para nuestro trabajo en este ensayo. Sostuvieron que todas las teorías prescriptivas deben considerar: (a) las cuestiones vinculadas con la metodología utilizada; (b) la manera en la que se juzgarán o se valorarán los datos; (c) el foco del usuario sobre el esfuerzo de evaluación. Las relaciones entre las teorías se presentan en un árbol de la teoría de la

evaluación, en el que cada una de las ramas principales representa una de las dimensiones: Métodos, Valoración y Uso. Cada teórico fue colocado en una de ellas y representa el principal énfasis que le pone a esa rama respecto a las tres. El sistema de categorización del árbol de la evaluación se basa en el énfasis relativo en los diversos modelos. Básicamente, cuando los evaluadores deben hacer concesiones, ¿qué es lo que rescinden con mayor facilidad y qué es lo que defienden con mayor tenacidad?

En la rama central del árbol de la teoría de la evaluación (ver Figura 1), la de los Métodos, la evaluación está guiada ante todo por la metodología de la investigación. Si bien la mayoría de los teóricos de la evaluación tienen preocupaciones metodológicas y ven a la investigación aplicada como la génesis de la evaluación de programas, algunos teóricos enfatizan firmemente esta orientación. Por lo general, los teóricos de esta rama están preocupados por la construcción de conocimiento (Shadish *et al.*, 1991) con la mayor rigurosidad posible, dadas las restricciones particulares dentro de esta evaluación. Para estas teorías, resulta fundamental el trabajo de Donald Campbell (1957), y posteriormente el de Campbell y Stanley (1966), que definen las condiciones para los diseños experimentales y cuasiexperimentales adecuados. Los escritos de la mayoría de los teóricos de esta rama se fundamentan en este trabajo o lo amplían.

A la derecha de la rama de los Métodos, nos encontramos con la rama de la Valoración, que inicialmente se inspiró en el trabajo de Michael Scriven. Fue este autor quien declaró que la evaluación no es tal sin la valoración (1967). Sostiene que el trabajo de la evaluación es realizar un juicio de valor sobre el objeto de estudio. Los teóricos de esta rama también se vieron influenciados por los escritos de Robert Stake (1967, 1975). Esta rama se divide en dos partes —la objetivista y la subjetivista— que distinguen las dos perspectivas fundamentales que informan el proceso de evaluación. La subrama influenciada por la parte objetivista, que se encuentra a la izquierda, se ve más fuertemente influida por las opiniones de Scriven. Considera que la función del evaluador es realizar esa valoración. Básicamente, el papel importante en materia de valoración le pertenece al evaluador. En la subrama subjetivista, que se encuentra a la derecha, los proponentes sostienen que la realidad es un proceso dinámico constante y que una verdad siempre es relativa a cierto marco de referencia particular.

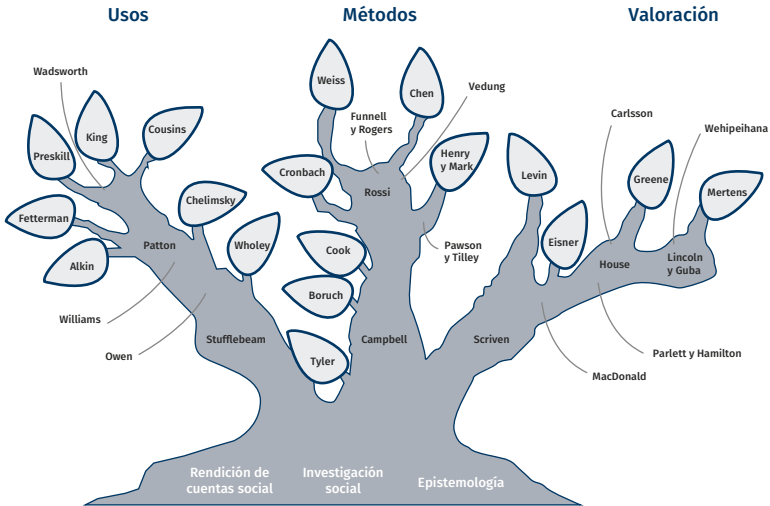
De este modo, la valoración debe darse en el contexto de entender la «importancia subjetiva» de la información de la evaluación. A diferencia de los objetivistas, estos teóricos no realizan solos y por sí mismos la decisión de valoración.

La tercera rama del árbol es el Uso que, en un principio, se centró en una orientación hacia la evaluación y la toma de decisiones. Básicamente, el trabajo de los primeros teóricos de esta rama se concentró específicamente en los que estaban empoderados para usar la información —habitualmente, las personas que habían contratado la evaluación—. Los teóricos posteriores expandieron la utilización a una audiencia de usuarios más amplia y a la construcción de capacidad de evaluación en la organización que se evalúa.

Las tres ramas del árbol no deben considerarse como independientes entre sí, sino que se han dibujado de manera que se refleje la relación entre ellas. Así, si el árbol estuviera en un espacio tridimensional, el extremo derecho de la rama de la valoración se relacionaría (o, tal vez, se tocaría) con el extremo izquierdo de la rama del uso. Es importante observar esto porque los teóricos se ubican en cada rama reflejando no solo su principal énfasis, sino también la otra tendencia importante de su enfoque en particular. De este modo, la naturaleza relacional de las tres ramas ayuda a representar mejor los enfoques de los teóricos.

Aquí presentamos la figura condensada para América del Norte, Europa, Australia y Nueva Zelanda (ver Figura 1). Puede resultar útil como marco de referencia para la discusión sobre la evaluación en los PIBM.

Figura 1. Árbol de la teoría de la evaluación



Fuente: Adaptado de Alkin, M. C. (2012). *Evaluation roots*. Segunda Edición. Thousand Oaks, California: Sage.

Los autores de la evaluación en el mundo en desarrollo: una visión general

En el presente trabajo, abordamos la cuestión de las perspectivas de la evaluación en los PIBM, a veces llamados «el mundo en desarrollo» o el «Sur global». Primero, las raíces de la evaluación en los PIBM, en un principio, fueron las establecidas por las organizaciones de asistencia del Norte. En este sentido, las raíces son institucionales más que independientes y contextualmente aseguradas por una perspectiva arraigada en métodos, valoración y uso. Los enfoques impuestos en este proceso se desarrollaron en América del Norte y en Europa, y, de ese modo, se basaron en el desarrollo en el campo de la evaluación en esas regiones más que en las tradiciones de evaluación de los PIBM.

A medida que la evaluación se convirtió en algo ampliamente utilizada en los PIBM y más localizada, las prescripciones en esta materia comenzaron a reflejar los contextos locales. Reiteramos, estas prescripciones de evaluación (o «teorías» descriptivas) fueron principalmente colectivas y no individuales. En otras palabras, ante todo fueron desarrolladas por

grupos de personas y no por autores individuales. Es tanto cultural como político en su origen. Mientras para la mayoría no es posible mencionar específicamente teóricos individuales como se hace en el libro «Las raíces de la evaluación» (Alkin, 2012), resulta importante encontrar un modo de destacar los esfuerzos constantes en el campo de la evaluación en el Sur. Así, comenzaremos con un marco de la evaluación basado en sus orígenes de uso en los PIBM y después daremos paso a una breve discusión de sus vínculos con el árbol de la teoría.

Esta perspectiva analítica de enfoques es una importante contribución no solo para la futura potencial identificación de teóricos para el árbol de la evaluación, sino también como fuente de presión para el cambio en el modo en el que el mundo en desarrollo se ocupa de la construcción del campo de la evaluación. Como afirmó uno de los autores, constituir el campo de la evaluación es fundamental para el uso eficaz de la evaluación en los PIBM (Carden, 2007). Es en la formación de este campo que los teóricos (ya sean colectivos o individuales) comenzarán a tener un lugar y una presencia más fuertes en la comunidad global de la evaluación.

Al elaborar este artículo, somos sumamente conscientes de que no representamos a las circunscripciones de la evaluación en los PIBM y esperamos que este ensayo estimule el pensamiento y los escritos de los evaluadores de los PIBM que quieran contribuir con la construcción de la práctica de evaluación (y potencialmente teoría) en sus países y regiones. También observamos que debido a que gran parte del trabajo de la evaluación del desarrollo que surge en los PIBM tiene una base local, no es conocido fuera de sus regiones de aplicación. En consecuencia, el presente ensayo está necesariamente incompleto, por lo que debería considerarse un marco para el pensamiento sobre la evaluación del desarrollo y no una recopilación completa de los metodólogos o de los enfoques metodológicos del Sur global. Lo que sugiere es la necesidad de que los analistas de la evaluación construyan un conocimiento más comprehensivo y una documentación completa sobre la evaluación del desarrollo, y, en términos más generales, sobre la construcción del campo de la evaluación en los PIBM. Queda claro que muchos científicos sociales de los PIBM que trabajan más ampliamente en metodología son activos practicantes de la evaluación y llevan las tradiciones metodológicas al campo de la evaluación, pero no

las documentan en el uso de la evaluación (por ejemplo, Chilisa, 2012). Mucho podría desarrollarse para fortalecer la práctica de evaluación en los PIBM a través de la definición y articulación del trabajo que se realiza de manera informal. Este es un componente esencial para la construcción del campo de evaluación en los PIBM.

Mientras que la evaluación existe en todos lados y de muchas formas, aunque como actividad informal, la evaluación del desarrollo se originó, en gran medida, en las agencias bilaterales de asistencia y en sus contrapartes multilaterales. La evaluación fue desarrollada y expandida por estas agencias como una herramienta para respaldar el traspaso de sus proyectos a los países en desarrollo. A medida que la asistencia comenzó a desarrollarse como una industria tras la independencia de muchos Estados del tercer mundo (en gran parte, en la década de los sesenta), rápidamente se hizo evidente la necesidad de algún tipo de respaldo sobre la base del modelo Bretton Woods para la reconstrucción de posguerra y que sería apoyado por los países de América del Norte y de Europa. Pronto se puso de manifiesto que se necesitarían algunas herramientas para valorar ese flujo de ayuda. Esto dio lugar al primer tipo de evaluación del desarrollo, al que llamaremos «metodologías adoptadas».

Metodologías adoptadas

Como su nombre lo indica, las metodologías adoptadas son aquellas que son transferidas de los sistemas desarrollados en América del Norte y en Europa para ser utilizadas en los PIBM, en especial, en la valoración de los flujos de asistencia. Los primeros esfuerzos en materia de evaluación en el Sur fueron considerados, en gran medida, una transferencia de tecnología mediante la actividad de capacitación en los enfoques particulares de evaluación que se usaban en las agencias bilaterales y multilaterales. El mayor propósito fue asegurar un cuerpo de evaluadores que podría satisfacer las necesidades de evaluación de las agencias que reconocían el valor de los consultores locales que comprendían tanto la cultura local como las reglas políticas y el idioma (¡acá estamos simplificando demasiado, pero no por mucho!). Uno de los autores discutió esta cuestión en otros artículos (Carden, 2007; Carden, 2010).

El análisis del marco lógico es una de las metodologías adoptadas clave que muchas agencias bilaterales y multilaterales pusieron en marcha. Evolucionó con el tiempo y, por lo general, se la llama «gestión basada en resultados». Gran parte de la capacitación en evaluación que realizan las agencias donantes se construye alrededor de enfoques para el análisis del marco lógico, o gestión basada en resultados, como se utiliza en las agencias bilaterales y multilaterales de asistencia para el desarrollo. El camino hacia los resultados (Morra-Imas y Rist, 2009) es una de las articulaciones más minuciosas de este enfoque y sirve como manual para los talleres de evaluación del desarrollo, como el Programa Internacional para la Capacitación en Evaluación para el Desarrollo (patrocinado por el Banco Mundial y la Universidad de Carleton). El enfoque también es duramente criticado (Gasper, 2000, Oficina de Servicios de Supervisión Interna de la ONU, 2008). Si bien Morra-Imas y Rist desarrollan y codifican los enfoques, el origen y la motivación del uso son fundamentalmente institucionales. Caería en la rama de uso del árbol de la teoría de la evaluación con su fuerte orientación al uso en la toma de decisiones por parte de quienes están empoderados para determinar el empleo de los fondos para el desarrollo.

Las agencias también prescriben otros métodos. Últimamente, los métodos experimentales y cuasiexperimentales han ganado mayor popularidad, en especial, en el pedido de más ensayos aleatorizados controlados (RCT, por su sigla en inglés). En su estudio de 2006, el Centro para el Desarrollo Global hizo un fuerte llamamiento para que hubiera más RCT en la evaluación del desarrollo como un modo de fortalecer la comprensión de qué funciona en materia de desarrollo (Grupo de Trabajo de la Brecha de Evaluación, 2006). Por supuesto, los RCT habían estado presentes durante largo tiempo, pero este estudio desató un ajetreo de actividad y de debate en la comunidad de evaluación del desarrollo y estimuló una considerable demanda de las agencias de más evaluaciones experimentales en materia de desarrollo. Gran parte del debate se centró en la relevancia de los RCT para algunas actividades de desarrollo. Si bien los proponentes de los RCT reconocen sus fortalezas y sus limitaciones, las agencias que establecen este enfoque como un requisito institucional ignoran este debate metodológico fundamental.

El estudio de 2006 realizado por el Centro para el Desarrollo Global condujo a la creación de la Iniciativa Internacional de Evaluación de Impacto (3ie, por su sigla en inglés). El mandato de la 3ie es respaldar los estudios de evaluación de impacto de calidad que pueden influenciar políticas y afectar muchas vidas. La iniciativa 3ie procura financiar estudios que produzcan nueva evidencia sobre qué funciona en el desarrollo y difundirla. La iniciativa 3ie apoya y fomenta un enfoque basado en la evidencia para la toma de decisiones y trabaja cada vez más para crear capacidad tanto para producir como para usar evaluaciones de impacto (sitio web de 3ie). Sus miembros incluyen quince agencias para el desarrollo (bilaterales, multilaterales y fundaciones) y cuatro agencias gubernamentales. La membresía se basa, en líneas generales, en una variedad de agencias ubicadas tanto en el Norte global como en el Sur global. La iniciativa 3ie lleva adelante gran parte de su trabajo por medio de pedidos de propuestas que son cuidadosamente analizados. Esta iniciativa también patrocina la *Journal for Development Effectiveness* (Revista para la Eficacia del Desarrollo). Esta revista publica los resultados de los estudios de impacto y también fomenta artículos sobre métodos de evaluación de impacto. La iniciativa 3ie respalda activamente los estudios con una clara hipótesis contrafáctica y que no se centran en un único método. En la actualidad, esta iniciativa concentra algunos de sus esfuerzos en la identificación de métodos de «N pequeña» que podrían usarse en la evaluación de impacto. La iniciativa 3ie se ha convertido rápidamente en una influencia importante para la evaluación del desarrollo. Su trabajo se incluiría en la rama de los métodos del árbol de la teoría de la evaluación, aunque con una tendencia hacia la rama del uso.

Las metodologías adoptadas tienen en común el hecho de que son una forma de transferencia de tecnología que, para bien o para mal, no ve una necesidad de adaptarse al contexto local. Otro aspecto en común es que tienden ser presentadas como un enfoque singular para los esfuerzos de evaluación de una agencia, por lo que se realiza todo esfuerzo posible para usar la metodología prescrita para todas las actividades. Además, representan una perspectiva organizacional y no la perspectiva singular de un metodólogo. Por último, sus orígenes son las necesidades de rendición de cuentas de las agencias más que las necesidades de aprendizaje de rendición de cuentas de aquellos a los que se apoya en el proceso de asistencia para el desarrollo. En suma, tienen un origen institucional.

Metodologías adaptadas

En parte como respuesta al papel esencialmente colonizador de la evaluación y al descontento con él, se le presta mayor atención al desarrollo de métodos que comprometen la participación de los profesionales de los PIBM —metodologías adaptadas—. Esta insatisfacción con los enfoques impulsados por los donantes, el reconocimiento de la importancia central del compromiso con los agentes de desarrollo de los PIBM y el reconocimiento de muchos de la necesidad de contar con diversos métodos para abordar diferentes cuestiones dio lugar a un enfoque distinto del diseño y de la aplicación de las metodologías de evaluación en algunos lugares. Muchas de estas metodologías todavía son desarrolladas, en gran medida por evaluadores y teóricos del Norte global, pero ellos están explícitamente preocupados por su aplicación en los PIBM y por la necesidad de abordar, con socios de los PIBM, el modo de adaptar el enfoque a diferentes escenarios socioculturales, políticos, económicos y ecológicos. Comienzan con el supuesto de que la metodología es sensible al contexto y no neutral a este. A diferencia de los métodos adoptados que se resumieron antes, tienden a agruparse en la rama del uso con una tendencia hacia la rama de la valoración.

Las metodologías adaptadas incluyen muchas metodologías participativas que fueron tipificadas en el trabajo de Robert Chambers (2008) como evaluación rural rápida y evaluación rural participativa. La primera surgió de la práctica, en particular en Tailandia, dada la necesidad de información rápida, oportuna, relevante, precisa y utilizable por parte de los responsables de tomar las decisiones. Era un contexto en el que gran parte de la información disponible era engañosa, pero también en el que no se podían completar encuestas amplias y detalladas de manera oportuna. La evaluación rural participativa «busca e incorpora formas participativas de empoderar a las personas locales y subordinadas, lo que les permite expresar y mejorar su conocimiento, y actuar» (Chambers, 2008: 85). Los procedimientos proporcionan una oportunidad para que las personas locales puedan tener un papel en la determinación del valor. Si bien Chambers no inventó estos métodos, su perspectiva sobre la investigación participativa y de acción fue instrumental en la formalización y la difusión de estos enfoques. Muchos otros —demasiados como para enumerarlos en este trabajo—, también se involucraron y fueron centrales en los desarrollos

de los métodos participativos. La cuestión aquí es que la formalización de los métodos ha estado en manos de los investigadores del Norte global, pero la articulación y la aplicación casi siempre en los PIBM, y fueron sensibles a los contextos en juego en los PIBM. La adaptación ha sido fundamental para el diseño. Estos métodos se sitúan en la rama de la valoración.

Otras metodologías adaptadas surgieron a partir de preocupaciones por las brechas metodológicas en la evaluación del desarrollo. A continuación, mencionaremos dos ejemplos, pero hay otros que también se ajustan bien a esta categoría. Las restricciones de espacio nos limitan a realizar un muestreo de enfoques.

Al primer ejemplo, el mapeo de resultados (Earl, Carden y Smutylo, 2001), le concierne explícitamente intentar abordar el enigma de la financiación del desarrollo: ¿cómo se valora la contribución al desarrollo cuando no se trata de una intervención singular que pueda asignarse y probarse aleatoriamente? ¿Para qué un proyecto o un programa de investigación del desarrollo puede adjudicarse el crédito racional y justamente? El mapeo de resultados reconoce que el desarrollo es sobre personas que se relacionan entre ellas y con su entorno. En vez de centrarse en los productos de un programa de desarrollo, el mapeo de resultados se concentra en los cambios en el comportamiento, actividades y relaciones de las personas y de las organizaciones con las que el programa de desarrollo trabaja de modo directo. Es específico sobre la teoría del cambio que se utiliza: qué cambios esperan los actores destinatarios de un programa y qué estrategias usan para alcanzarlos. El mapeo de resultados proporciona un marco para la reflexión, a medida que el programa evoluciona, y de ese modo, mejora el potencial de uso. Fue adoptado en una amplia gama de programas y fue adaptado de varias formas. Por ejemplo, Deprez, Nirarita y Shatifan (2010) confeccionaron una guía acerca del uso del mapeo de resultados para los programas de desarrollo de la comunidad en Indonesia, a partir de su experiencia de campo en ese país, e hicieron las adaptaciones que conservaban las bases teóricas y las raíces filosóficas del mapeo de resultados, pero adaptaron el lenguaje y el enfoque a las necesidades y al estilo de sus programas en Indonesia. Asimismo, Loza (2010) llevó adelante el trabajo de mapeo de resultados con un programa muy diferente, al investigar sobre las cuestiones comerciales en América Latina y escribió

sus experiencias en ese contexto, de nuevo, con adaptaciones al tema y al contexto geográfico. El mapeo de resultados se ubica en la rama del uso del árbol de la teoría de la evaluación.

Otra metodología adaptada es el «Cambio más significativo». Fue diseñada en 1996 por Rick Davies y Jess Dart (Davies y Dart, 2005) mientras trabajan en el sur de Asia. Si bien es una forma de monitoreo participativo y evaluación, es notable por su importante aceptación y por sus orígenes en el campo. Davies y Dart bregaban para determinar de qué modo evaluar programas grandes con múltiples actores sin indicadores claros. A medida que ambos se enfrentaban con el mismo problema en proyectos separados, dialogaban y fueron construyendo el enfoque que Davies desarrolló en su trabajo en Bangladesh. Este se basa en las historias de cambio significativo en relación con un programa. El análisis se realiza con los participantes del proyecto y los hallazgos son utilizados para monitorear el impacto y para basarse en los éxitos. El cambio más significativo integra al usuario directamente a un proceso de evaluación. Se identifican las historias de cambio significativo en el campo y se comparten con un comité de selección (del programa); este elige la historia más significativa de las contadas para compartirla con el siguiente nivel de la organización (podría ser un órgano regional o nacional); a su vez, este órgano selecciona las historias más significativas para compartirlas con otros —por ejemplo, donantes, público interesado y agencias gubernamentales—. Los propios actores analizan las historias de cambio y las eligen por la representatividad y por la importancia de compartirlas con otros. Están orientadas no solo a compartir el éxito, sino también a aprender de lo sucedido —el cambio significativo puede ser negativo— y a mejorar para el futuro. El objetivo de este enfoque es ayudar a organizar una enorme cantidad de información diversa que todos relacionan con el progreso del programa. Su naturaleza hace que sea útil para identificar tanto resultados inesperados como esperados. La variedad apoya el desarrollo de una rica imagen de un complejo proceso de cambio. La metodología del cambio más significativo se sitúa en la rama del uso del árbol de la teoría de la Evaluación.

La evaluación del desarrollo, según como es descrita por Patton (2010), claramente también encaja aquí. Como la define Michael Quinn Patton, «la evaluación del desarrollo respalda el desarrollo de la innovación para

guiar la adaptación a realidades emergentes y dinámicas en entornos complejos» (2010: 1). Coloca al evaluador dentro del programa al apoyar el pensamiento evaluativo y la reflexión permanente para ayudar a los programas a manejar entornos complejos y cambiantes. En el diseño de la evaluación del desarrollo, Patton se vio influenciado por su trabajo no solo en América del Norte, sino también con las comunidades de pueblos originarios. Sin duda, la evaluación del desarrollo se encuentra en la rama del uso, reflejado por el hecho de que Patton ya estaba presente en esta rama, pero la evaluación del desarrollo como enfoque se situaría en un punto más alto de la rama del que Patton ocupa hoy en el árbol.

Varias otras metodologías, como la evaluación realista (Pawson y Tilley, 1997) y el «dar sentido» (Snowden, 2002), se adaptan activamente para ser usadas en la evaluación del desarrollo. No están incluidas en este artículo porque fueron desarrolladas para otros escenarios, no para abordar las brechas de la evaluación del desarrollo. Una tipología más detallada podría encontrarles un lugar en esta discusión.

La característica común de las metodologías adaptadas es que, mientras se formalizan en el Norte global, incluyen un compromiso importante con el Sur global y, en algunos casos, se basan directamente en la práctica de esta región. Asumen una especificidad de contexto y esperan modificaciones de la metodología.

Metodologías autóctonas

Identificamos algunas metodologías de evaluación desarrolladas en el Sur global. Las distinciones no siempre están bien definidas. A modo de ejemplo, la evaluación rural rápida tiene fuertes raíces en Tailandia, pero elegimos identificarla como una metodología adaptada debido al fuerte papel de las personas y las organizaciones de Europa en la formalización del método. Las metodologías autóctonas parecen ser específicas de una región y, en general, no se comparten efectivamente a través de las regiones. Algunas se relacionan con los esfuerzos del Norte global, pero sus diseños y raíces están vinculados con la teoría y la práctica locales. De hecho, cierto desarrollo de los métodos autóctonos de evaluación también tiene lugar en el Norte global (ver, por ejemplo, el capítulo 30 de

la segunda edición de «Las raíces de la evaluación» para un debate sobre un trabajo que emplea conceptos maoríes desarrollados por Wehipeihana, McKegg, Cram y otros). En este artículo, resumimos tres ejemplos: uno de África, otro de Asia del Sur y otro de América Latina. Sin duda, existen otros; los que mencionamos son ilustrativos de las aplicaciones en uso. Como Alkin, Christie y Vo observan en el capítulo 30 de «Las raíces de la evaluación» (2012), la disposición de los métodos en el árbol de la teoría de la evaluación puede ser desafiante en extremo. Si bien cada uno de los ejemplos que se mencionan a continuación es único, todos reflejan la estrecha interacción entre las ramas de la valoración y del uso del árbol de la teoría de la evaluación y, a menudo, expresan una brecha en el abordaje de los valores locales en las metodologías desarrolladas en los países industriales que los reflejan más estrechamente.

El Mecanismo Africano de Revisión entre Pares (APRM, por su sigla en inglés, desarrollado por la NEPAD —Nueva Alianza para el Desarrollo de África—) es un buen ejemplo. Este se desarrolla, evoluciona y se estudia de forma autóctona. Es un conjunto específico de procedimientos para la autoevaluación de los países en materia de gobernanza y de derechos humanos. Si bien el enfoque se basa en el Mecanismo de Revisión entre Pares de la OCDE [Organización para la Cooperación y el Desarrollo Económicos] (OECD, 1997), fue desarrollado en el marco de la NEPAD por investigadores y responsables de formular políticas africanos y es aplicado por equipos autóctonos y conduce a evaluaciones basadas en África (entre otros Kanbur; 2004; Herbert y Gruzd, 2008; Mbelle, 2010; Turianskyi, 2010). Casi la mitad de los países de África se sumaron a este esquema voluntario que se enfoca en la gobernanza (política, económica y empresarial) y en el desarrollo socioeconómico en su revisión entre pares.

Un mecanismo de revisión entre pares es, como su nombre lo indica, un proceso formal por el que los pares de una persona revisan su política y su práctica en un campo como la gobernanza, la política del desarrollo, entre otros, y le proveen retroalimentación basada en un conjunto de principios y refleja la experiencia de esa persona y las observaciones sobre la mejora que podrían realizar en su enfoque. La OCDE introdujo un sistema de revisión entre pares hace más de cuarenta años, el cual es muy usado en esta organización, y fue adoptado y adaptado en muchos otros

escenarios. El mecanismo africano es uno de los más significativos y aquí lo consideramos autóctono debido al importante compromiso de los agentes y agencias africanas en su diseño, implementación, crítica y evolución. Es un proceso voluntario y se basa en el informe de autoevaluación de un país. La autoevaluación se desarrolla como un diálogo nacional que está abierto y a la espera de aportes de todos los sectores, incluidos el no gubernamental y el privado. La revisión incluye no solo una misión de evaluación, sino también un seguimiento a un Plan Nacional de Acción para lograr cambios inmediatos y a más largo plazo que se proponen sobre la base de los hallazgos de la revisión entre pares. La revisión entre colegas de las evaluaciones nacionales es realizada por los jefes de Estado y por los funcionarios sénior. A veces, los procesos son criticados, como la revisión entre pares de Sudáfrica (Mbelle, 2010; Turianskyi, 2010), y algunas de estas críticas alcanzan a los procesos continentales del APRM. Esta es una señal en sí misma significativa de la importancia que se le da al proceso y a la que le concierne garantizar el progreso de un proceso relativamente joven y sin experiencia. Un comentario sobre el mecanismo realizado por Cilliers en 2003 continúa siendo relevante hoy: «Quedan muchas preguntas (...) el debate sobre la revisión entre pares abrió un espacio para que las organizaciones de la sociedad civil de África busquen y establezcan procesos análogos para que los gobiernos y los líderes africanos rindan cuenta de los compromisos asumidos y de sus decisiones. Quizá, es aquí donde se podría encontrar la mayor esperanza de que haya una rendición de cuentas y una revisión eficaz» (Cilliers, 2003: 14). El APRM se encuentra en la rama de uso del árbol de la teoría de la evaluación con una inclinación hacia la valoración.

La sistematización (o *systematization* en inglés) es una metodología latinoamericana desarrollada por un pequeño equipo a finales de la década de los ochenta. A diferencia del APRM, la sistematización se sitúa en la rama de la valoración con una inclinación hacia la del uso. Se trata de un enfoque que tiene sus raíces en el trabajo de Paulo Freire y en la investigación-acción participativa, aunque con el tiempo se adaptó al contexto latinoamericano. Como sucede con muchos métodos autóctonos y con gran parte de la evaluación del desarrollo en términos más generales, las raíces del diseño son múltiples, por lo que será extremadamente difícil atribuírselas a una persona (Oscar Jara, Sergio Martinic, Daniel Selener,

Mariluz Morgan y Esteban Tapella serían miembros de un grupo central significativo). Mucha de la documentación sobre la metodología en inglés (León, sin fecha) fue respaldada por ActionAid, agencia británica no gubernamental para el desarrollo con activa participación en América Latina. Los autores antes mencionados continúan publicando en América Latina sobre la experiencia con la sistematización (Tapella, 2011).

La sistematización es el «proceso participativo y serio de reflexión acerca de los distintos aspectos de un proyecto específico o de la intervención general: actores, su interacción, productos, resultados, impacto, problemas y procesos» (Tapella, sin fecha: 24). El concepto básico es que los participantes reconstruyan, analicen e interpreten una experiencia con el fin de aprender para futuras intervenciones. Por lo general, se emplea como un proceso constante de reflexión y de aprendizaje en un proyecto o intervención. Como sucede con el cambio más significativo y con el mapeo de resultados, la sistematización tiende a apoyar la explicación de los resultados y los impactos inesperados. Se basa en el supuesto de que muchas intervenciones no suceden como se planifican, por lo que una documentación sistemática y una reflexión formal y documentada sobre la real trayectoria que toma una intervención ayuda tanto a identificar ese camino como a aprender para mejorar.

En un artículo breve, Hansford, Santana y Reinoso (2007) describen una experiencia usando la sistematización con una red de organizaciones no gubernamentales basadas en la comunidad enfocadas en el fortalecimiento de la seguridad alimentaria en América Latina. La red surgió de una serie de proyectos de un grupo que quería conservar la reflexión y el aprendizaje, que habían sido importantes para el éxito de los proyectos, para futuras iniciativas. Con el apoyo de uno de los donantes internacionales, se capacitó a un núcleo de personas como facilitadores en sistematización. La confianza que esto creó entre los profesionales del desarrollo les dio el impulso tanto para mejorar los programas como para difundir su experiencia. En un caso, la Asociación para el Desarrollo Sostenible y Solidario de la región sisalera en Brasil, los elementos clave de los programas de microfinanzas y de asistencia técnica que se desarrollaron se incorporaron a las políticas nacionales para fortalecer la agricultura a pequeña escala.

En India, el Centro de Asuntos Públicos implementó la Tarjeta de Calificación Ciudadana (ver <http://www.citizenreportcard.com/>). Fue diseñada para que los ciudadanos pudieran dar retroalimentación a las agencias públicas, teniendo en cuenta el uso local en el contexto de ese país. La Tarjeta de Calificación Ciudadana se encuentra en la rama del uso del árbol de la teoría de la evaluación, con una fuerte orientación hacia la rama de la valoración debido a la importancia que le da al empoderamiento ciudadano en su diseño y en sus procesos.

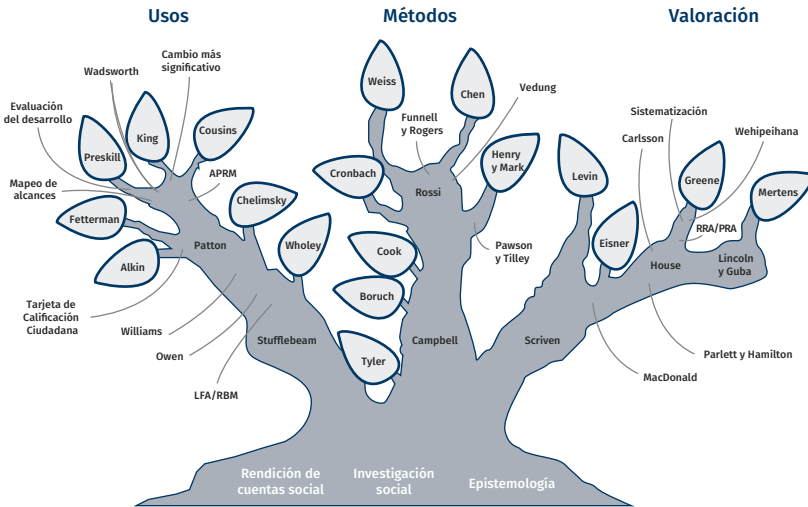
Por medio de encuestas participativas, la Tarjeta de Calificación Ciudadana se propone mejorar el acceso y provisión de los servicios públicos. La documentación disponible es instrumental, los informes se encuentran en los sitios web del proyecto (como <http://www.adb.org/projects/e-toolkit/Naawan-Citizen-Report.pdf>) y en los sitios de intercambio de conocimiento, como Civicus (http://pgexchange.org/index.php?option=com_content&view=article&id=164&Itemid=160). Algunos informes fueron formalmente publicados (Sekhar, Nair y Prabhakar, 2008). Este documento informa el ranking de calidad de los servicios públicos brindados por los Gram Panchayats (órganos locales de gobierno) de dos distritos del estado de Chattisgarh (India). Las conclusiones se debatieron abiertamente en los distritos, pero no hay ningún dato de seguimiento fácilmente disponible sobre si se realizaron o no cambios sostenibles desde que se reportó el trabajo.

En la India, el Centro de Asuntos Públicos (<http://www.pacindia.org>) conserva escasa documentación sobre las Tarjetas de Calificación Ciudadana, pero, como sucede con muchas agencias orientadas a la acción, participa en la exclusiva publicación académica de su trabajo. Este es un enfoque importante desarrollado de manera autóctona que es muy conocido en Asia del Sur y en el pequeño núcleo de personas con fuerte interés en las encuestas basadas en la comunidad. Merece más atención en la comunidad de investigación de la evaluación. Por ejemplo, sería importante documentar el seguimiento de dónde se presentaron las Tarjetas de Calificación Ciudadana (como en el caso anterior) y, a través de ese proceso, mejorar potencialmente las tarjetas de calificación y su uso. Esto permitiría ampliar el alcance de la experiencia y el uso del enfoque más allá de la experiencia del Centro de Asuntos Públicos.

Ubicación en el árbol de la teoría de la evaluación

Claramente, no hemos encontrado un fuerte enfoque de los evaluadores del Sur como teóricos individuales. No obstante, existen influencias sistemáticas en la evaluación por medio de organizaciones e instituciones. Todo esto es interesante, y creemos que vale la pena seguir de alguna forma, aunque no se ajuste tan cuidadosamente a la perspectiva analítica con la que intentábamos trabajar en la vinculación con el libro «Las raíces de la evaluación». Todavía observamos conexiones y debido a la necesidad de intentar mostrar una relación con la teoría de la evaluación en el Norte, deberemos ser complacientes. Evidentemente, se pueden incluir algunos métodos autóctonos. En ese sentido, en el árbol de la teoría de la evaluación, intentaremos colocar contribuciones al desarrollo de la teoría que fueron desplegadas (como planificadas) en grupo, con algunos de los nombres de quienes se identifican con el grupo señalado en el texto (ver Figura 2). Asimismo, las adaptaciones a los procedimientos de evaluación del Norte (como los hemos llamado) serán incluidas cuando muestren ser prometedoras para potencialmente evolucionar en una única teoría que refleje los valores y los contextos locales. De nuevo, así planificadas.

Figura 2. Árbol modificado de la Teoría de la Evaluación



Fuente: Adaptado de Alkin, M. C. (2012). *Evaluation roots*. Segunda Edición. Thousand Oaks, California: Sage.

En el Cuadro 1, presentamos los enfoques de los PIBM que se encuentran en la Figura 2 en columnas para cada una de las ramas del árbol e indicamos cuál de los enfoques identificados en los PIBM se encontrará en cada una. Además, estos enfoques se colocaron en el cuadro de una manera que se corresponde aproximadamente con su ubicación en la rama del árbol (de arriba abajo). Cuando fue posible, se especificaron las organizaciones o las personas que se identifican con un enfoque. Por último, como recordatorio, señalamos si el enfoque es adoptado, adaptado o autóctono.

Cuadro 1. Enfoques de la evaluación del desarrollo y el árbol de la teoría

| Uso | Métodos | Valoración |
|---|---|--|
| Evaluación del desarrollo <i>Patton</i> (Adaptado) | | Evaluación rural rápida, evaluación rural participativa <i>Chambers</i> (Adaptado) |
| Mapeo de alcances <i>Earl, Carden y Smutyló</i> (Adaptado) | | Sistematización <i>Selener, Tapella et al.</i> (Autóctono) |
| Cambio más significativo <i>Davies y Dart</i> (Adaptado) | 3ie: Iniciativa Internacional para la Evaluación del Impacto <i>Centro para el Desarrollo Global</i> (Adoptado) | |
| Tarjeta de Calificación Ciudadana <i>Centro de Asuntos Públicos, India</i> (Autóctono) | | |
| Mecanismo Africano de Revisión de Pares <i>Nueva Alianza para el Desarrollo de África</i> (Autóctono) | | |
| Análisis del marco lógico/Gestión basada en resultados <i>Muchas agencias bilaterales para el desarrollo, así como bancos multilaterales de desarrollo (por ejemplo, el Banco Mundial)</i> (Adoptado) | | |

Comentario al margen

Una consecuencia no prevista de esta investigación es la preocupación sobre la aplicabilidad de las ramas del árbol al considerar la evaluación en los PIBM. Como se señaló antes, existen relaciones entre las ramas del árbol de la teoría de la evaluación que tienen cierta conexión entre las ramas del uso y de la valoración —no son extremos opuestos, sino que se inclinan una hacia la otra—. Sobre la base de su experiencia en los PIBM, uno de los autores (Carden) propone aquí una nueva rama en crecimiento que abarca el uso y la valoración. A medida que los conceptos de sistemas se arraigan cada vez más en la evaluación y que un número creciente de autores abordan el pensamiento sistémico como una influencia clave en

el campo, la complejidad surge como una influencia impulsora sobre a qué sería más probable que renuncien algunas metodologías en cualquier compromiso. Ni el mapeo de resultados ni la evaluación del desarrollo se sitúan de manera totalmente cómoda en la rama del uso. Una influencia impulsora clave en estos enfoques (y en otros nuevos) es su foco de surgimiento como un factor clave. Se basan en la premisa de que, en general, trabajamos con cuestiones que no pueden saberse por anticipado. Por consiguiente, nuestro punto de partida no es un conjunto claro de objetivos y de acciones para una intervención, sino un resultado de alto nivel y los primeros pasos que se pueden adoptar para ir en esa dirección. Puesto que el contexto interviene a través de los cambios políticos, sociales, económicos y ecológicos, las variaciones de trayectoria y los programas deben modificarse necesariamente para dar cuenta de estos cambios.

Por un lado, el otro autor (Alkin) plantea que la complejidad puede no ser una rama nueva, sino solo una extensión de la rama del uso, es decir, una preocupación que se considera al buscar alcanzar el uso. De este modo, la complejidad como potencial característica definitoria se parece mucho más a otra etapa de la rama del uso que, en su base, observó la evaluación y la toma de decisiones, pero que, posteriormente, tuvo una definición más amplia para los usuarios en términos generales y para incrementar la construcción de la capacidad organizacional.

Por el otro, puede haber una necesidad, dada la diversidad del mundo de los PIBM, de crear una nueva rama denominada «contexto» que avale la complejidad y la incertidumbre en estos contextos como la característica definitoria. Obviamente, hay innumerables y muy diversos contextos que deben considerarse. Este autor reconoció la cuestión del contexto en su capítulo «Evaluación sensible al contexto» del libro «Las raíces de la evaluación» (Alkin, 2012).

Si bien los autores de esta obra reconocen que ambos son de América del Norte, reciben con beneplácito el aporte de quienes están más familiarizados con la evaluación en los PIBM. Esperamos que reflexionen sobre estas cavilaciones.

Notas

* Título original “*Evaluation Roots: An International Perspective*” en Journal of MultiDisciplinary Evaluation, Volumen 8, N.º 17, ISSN 1556-8180, Enero 2012. Publicación autorizada en inglés © 2012 en Journal of Multidisciplinary Evaluation. Traducción autorizada al español. Todos los derechos reservados.

** Nota de los autores: Las opiniones expresadas pertenecen a los autores. Los autores agradecen a Tarek Azzam por su ayuda con las figuras del árbol de la teoría de la evaluación.

1. Nota de la editora: sitio web no encontrado. Fecha de consulta: 28 de agosto de 2016.
2. Nota de la editora: sitio web no encontrado. Fecha de consulta: 28 de agosto de 2016.

Referencias

Alkin, M. C. (Ed.) (2004). *Evaluation roots: tracing theorists' views and influences*. Thousand Oaks, California: Sage.

— (2012). *Evaluation roots*. Segunda edición. Thousand Oaks, California: Sage.

Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.

Campbell, D. T. y Stanley J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, Illinois: Rand McNally.

Carden, F. (2007). The real evaluation gap. *Alliance*, 12(4), 53-54.

— (2010). Introduction to the Forum on Evaluation Field Building in South Asia. *American Journal of Evaluation*, 31(2), 219-221.

Center for Global Development. Evaluation Gap Working Group. (2006). *When will we ever learn: Improving lives through impact evaluation*. Washington, D. C.: Center for Global Development.

Chambers, R. (2008). *Revolutions in development inquiry*. Londres: Earthscan.

Chilisa, B. (2012). *Indigenous research methodologies*. Londres: Sage.

Cilliers, J. (2003). *Peace and security through good governance: a guide to the NEPAD African Peer Review Mechanism* (Ensayo N.º 70). Pretoria, Sudáfrica: Institute for Security Studies.

Davies, R. y Dart, J. (2005). *The 'Most Significant Change' technique: A guide to its use*. Disponible en <http://mande.co.uk/docs/MSCGuide.pdf>

Deprez, S., Nirarita, E. y Shatifan, N. (2010). *Outcome mapping: Jejak perubahan menuju kerberhasilan*. Denpasar: VECO Indonesia.

Earl, S., Carden, F. y Smutylo, T. (2001). *Outcome mapping: Building learning and reflection into development programs*. Ottawa: International Development Research Centre.

Gaspar, D. (2000). Evaluating the 'logical framework approach' towards learning-oriented development evaluation. *Public Administration and Development*, 20, 17-28.

Hansford, F., Araujo, V., Santana, L. y Reinoso, G. H. (2007). From participatory systematization to a regional network for policy change. *Participatory Learning and Action*, 56, 33-40.

Herbert, R. y Gruzd, S. (2008). *The African Peer Review Mechanism: Lessons from the pioneers*. Johannesburg, Sudáfrica: South African Institute of International Affairs.

International Initiative for Impact Evaluation (3ie). (Sin fecha). Sitio web de la Iniciativa para la Evaluación de impacto. Disponible en <http://www.3ieimpact.org/>.

Kanbur, R. (2004). The African Peer Review Mechanism: An assessment of concept and design. *Politikon*, 31, 157-166.

León, R. (Sin fecha). Systematizing and learning from our experience on advocacy and campaign in the Americas. En R. León (Ed.), *Advocacy for change*. Londres: ActionAid.

Loza, J. (2010). *Monitoring and evaluating research outputs: The use of outcome mapping in the Latin American Trade Network*. Saarbrücken: Lambert Academic Publishing.

Marrow, A. (1969). *The practical theorist: The life and work of Kurt Lewin*. Nueva York, Nueva York: Basic Books.

Mbelle, N. (2010). *The APRM process in South Africa*. Rosebank, Sudáfrica: Open Society Initiative for Southern Africa.

Morra-Imas, L. y Rist, R. (2009). *The road to results: Designing and conducting effective development evaluations*. Washington, D. C.: The World Bank.

Organization for Economic Co-operation and Development (OECD) (1997). *Peer review: A tool for co-operation and change*. París: Organization for Economic Co-operation and Development.

Pawson, R. y Tilley, N. (1997). *Realistic evaluation*. Londres: Sage.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Nueva York, Nueva York: Guildford.

Scriven, M. (1967). The methodology of evaluation. En R. E. Stake (Ed.), *Curriculum evaluation* (Asociación Americana de Investigación Educativa, Serie Monográfica sobre Evaluación N.º 1). Chicago, Illinois: Rand McNally.

Sekhar, S., Nair, M. y Prabhakar, K. (2008). *Public services provided by Gram Panchayats in Chhattisgarh*. Chhattisgarh y Bangalore: Smarthan and Public Affairs Centre.

- Selener, D., Purdy, C. y Zapata, G. (1996). *A participatory systemization workbook: documenting, evaluating and learning from our development projects*. Quito, Ecuador: International Institute for Rural Reconstruction.
- Shadish, W. R., Cook, T. D. y Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, California: Sage.
- Snowden, D. (2002). Complex acts of knowing: Paradox and descriptive self-awareness. *Journal of Knowledge Management*, 6(2), 100-111.
- Stake, R. E. (1967). *The countenance of educational evaluation*. *Teachers College Record*, 68, 523-540.
- (1975). *Program evaluation particularly responsive evaluation* (Documento de divulgación N.º 5). Kalamazoo, Michigan: The Evaluation Center.
- Tapella, E. (2011). ¿Cómo aprender desde la práctica? Aproximaciones conceptuales y metodológicas para la sistematización de experiencias de desarrollo, en *Perspectivas en Políticas Públicas*, Vol. II, N.º 4 (en imprenta).
- (Sin fecha). Systematization: Basic concepts and methodological considerations. En R. León (Ed.), *Advocacy for change*. Londres: ActionAid.
- Turianskyi, Y. (2010). *Off track? Findings from South Africa's first APRM implementation report*. Johannesburgo, Sudáfrica: South African Institute of International Affairs.
- United Nations Office of Internal Oversight Services. (2008). *Review of results-based management at the United Nations*. Nueva York, Nueva York: United Nations.

Modelos teóricos para la práctica de la evaluación de programas*

Xavier Ballart

I. Introducción

¿Qué valor público añade a la sociedad la promoción pública de viviendas, la construcción de nuevas infraestructuras de transporte rápido o la inversión en protección medio ambiental? ¿Qué se obtiene de la inversión en I+D, de la subvención de los estudios universitarios o de un programa de guarderías para madres trabajadoras? ¿Qué cabe esperar de los servicios preventivos de seguridad, sanitarios o sociales?

La evaluación de los programas públicos —y de las políticas en que estos se fundamentan— es una disciplina científica que se justifica por la presión que experimentan las instituciones públicas por determinar en qué medida la intervención pública produce una mejora en el bienestar individual o social, cómo se produce esta mejora y cómo se podría conseguir de una forma más efectiva.

A diferencia del sector privado donde existe un criterio fundamental —el criterio del beneficio— para valorar la actividad de las personas y de las organizaciones, en el sector público, raramente existen unos criterios claros y ampliamente aceptados sobre los que fundamentar la valoración de la actividad de las instituciones político-administrativas. Intuitivamente podemos tener una cierta idea de si se produce igualdad en el acceso a determinado bien público, si una solución es económicamente eficiente, o sobre el nivel de pobreza en un determinado caso en relación con otros,

pero resulta muy difícil comparar, en términos de bienestar general, distintas «cantidades» de estos valores. La monetarización de este tipo de valores resulta muy difícil, lo que complica enormemente el análisis. La adopción de un criterio en perjuicio de otro se debe justificar, es preciso elaborar medidas del grado en que se consiguen estos objetivos, además de encontrar la manera en que se pueden introducir mejoras.

En la literatura sobre evaluación se señala su carácter aplicado y práctico en la medida que el objeto de estudio está determinado por los programas y las políticas de las instituciones públicas. Asimismo se señala el carácter de «arte», ante las limitaciones que encuentra una aproximación «científica» al producirse la investigación en un contexto real. Sin embargo, ello no significa que la evaluación de programas carezca de una base teórica que se ha ido desarrollando a partir de las teorías básicas de otras disciplinas y de la acumulación de experiencias en áreas sectoriales específicas. Sería un error infravalorar la importancia de la teoría en la evaluación de programas y desaprovechar las lecciones de la experiencia.

Este artículo se divide en dos partes. La primera parte es una síntesis de las aportaciones de siete autores fundamentales para el desarrollo teórico de la evaluación de programas sobre la base de la selección de Shadish, Cook y Leviton (1991) —en el mismo sentido, Shadish y Epstein en un trabajo previo (1987)—. Esta selección de autores permite resumir en pocas páginas la evolución histórica de la disciplina y los cambios que se han ido produciendo en el campo conceptual, introduciendo a su vez referencias a los trabajos de otros autores que elaboran más extensamente algunas de las ideas propuestas inicialmente por los autores seleccionados. En la segunda parte de este artículo, se elabora una categorización de las distintas dimensiones teóricas y prácticas que aparecen en las obras de estos autores con el objeto de explicar las tendencias en el desarrollo de la teoría de la evaluación y ofrecer un marco útil para encuadrar su ejercicio por académicos y/o consultores profesionales.

II. Los fundamentos teóricos

La evaluación de programas es una disciplina cuyo origen y desarrollo inicial se produce en Estados Unidos, fundamentalmente a raíz de la legislación

federal que obliga, y financia a la vez, la evaluación de los programas que impulsa. Esta es la base sobre la que se establece la profesión, lo que explica la variedad de enfoques y perspectivas. Los primeros programas cuya evaluación fue ordenada por la legislación federal fueron en educación si bien rápidamente este requisito se extendió a otros sectores: formación ocupacional, lucha contra la pobreza, delincuencia juvenil, vivienda.

Los siete autores seleccionados proceden de tradiciones científicas diversas y han desarrollado su profesión en departamentos universitarios, institutos o administraciones dedicados a distintas áreas: Campbell, Stake y Cronbach proceden de la psicología. Sin embargo, Stake, al igual que Weiss, ha desarrollado su carrera académica en departamentos de educación. En el caso de Weiss, socióloga de formación, su vinculación como investigadora o consultora con distintos programas de la administración federal norteamericana le dan una visión más amplia no circunscrita a temas de educación. Lo mismo cabe afirmar de Wholey, un matemático cuya carrera profesional se desarrolla fundamentalmente en la administración federal. Por último, Rossi es uno de los sociólogos más prestigiosos de los Estados Unidos mientras que Scriven es un filósofo de la ciencia¹.

1. Scriven. La ciencia de la valoración

Scriven es uno de los primeros teóricos de la evaluación. Su artículo *The methodology of evaluation* (1967, 1972) es uno de los artículos más citados en evaluación educativa (Smith, 1981 cit. por Shadish, Cook y Leviton, 1991). En este artículo introduce los términos *evaluación formativa* y *evaluación sumativa* para distinguir entre las evaluaciones cuya realización tiene por objetivo proporcionar información a las personas que intentan mejorar el objeto evaluado, de las evaluaciones pensadas para proporcionar información sobre si se debe continuar o terminar un programa. Scriven muestra una clara preferencia por la evaluación sumativa en la medida que la primera requiere un conocimiento que va más allá de si un programa ha funcionado mejor o peor, lo que no siempre es factible.

Scriven es el autor que más ha profundizado en la idea de evaluación como ciencia de la valoración. Critica a los que definen evaluación como producir información para la toma de decisiones. La «evaluación es lo

que es, la determinación de mérito o valor, para lo que sea utilizada es otra cuestión» (1980: 7). «La investigación evaluativa debe producir como conclusión exactamente el tipo de afirmación que a los científicos sociales se les ha dicho es ilegítima: un juicio de valor o mérito» (1974: 4).

Scriven entiende que la sociedad reclama una ciencia de la valoración porque la sociedad necesita saber si su administración, sus funcionarios y sus programas son buenos. Un programa será un buen programa si satisface las necesidades más importantes de la población. «La evaluación debería permitir determinar la necesidad que atiende el rendimiento actual del programa y los rendimientos que mejor darían satisfacción a grupos de necesidades» (1981: 130).

Para Scriven, la evaluación debería servir a los intereses de los afectados por el programa que está siendo evaluado. La evaluación no debe hacerse desde la perspectiva de la gestión del programa, ni tampoco desde la perspectiva del usuario o cliente. Se trata de una perspectiva que debería estar por encima de estos dos puntos de vista, algo que podríamos llamar «el interés general» y que, sin embargo, señala para cada una de las partes la importancia del punto de vista de las demás (1980: 103).

La evaluación de un programa no siempre está justificada: las evaluaciones deberían satisfacer el criterio de costo-eficacia: los beneficios fiscales de hacer una evaluación deben exceder los costos de su realización, en caso contrario no debería hacerse (1976: 220).

Scriven quiere minimizar la posibilidad de sesgo al formular un juicio de valor. Para ello identifica algunas razones por las que se tiende a sesgar la evaluación: el contexto organizativo puede llevar al evaluador a poner los intereses de la administración por encima de los demás. La lealtad al programa o a la institución que lo financia, la pertenencia como personal a la organización (evaluación interna), o la cooptación del evaluador por el programa convirtiéndolo en un defensor del mismo, son algunos de los elementos contextuales que pueden estar en el origen de este problema.

La tendencia a evaluar los programas en función de los objetivos de sus responsables y técnicos es la otra principal causa de sesgo en la valoración,

según Scriven. Para este autor, la evaluación debe servir para identificar todos los efectos, intencionados o no, que puedan ayudar a resolver los problemas sociales. Los grandes objetivos de un programa normalmente son vagos y abstractos, dado que lo que pretenden es obtener apoyo político, mientras que los objetivos que se fijan profesionales y técnicos para un programa también suelen estar sesgados dado que se seleccionan aquellos que es más factible conseguir. Por todo ello, Scriven es un claro partidario de la evaluación *goal free*, en función de las necesidades y no de los objetivos afirmados, Scriven recuerda que la justicia es ciega al igual que la medicina, que lo es doblemente. En evaluación, ni el paciente, ni el médico, ni el evaluador conocen de antemano qué efectos cabe esperar. «El evaluador tiene que descubrir los efectos» de un programa y comparar «los efectos con las necesidades de aquellos a los que afecta» (1983: 235).

Otras formas de controlar los posibles sesgos es haciendo circular los resultados de una evaluación a todas las partes afectadas, dando tiempo para su discusión. Esta ya es una forma de metaevaluación, otro término acuñado por Scriven para referirse a la necesidad de que las evaluaciones sean a su vez evaluadas. Para ello, Scriven utiliza una lista de cuestiones (objeto, cliente, contexto, recursos disponibles, función, beneficiarios, necesidades y valores, estándares, limitaciones, resultados, posibilidad de generalizar, costo...) que es preciso tener en cuenta antes de iniciar una evaluación (1980: 113-116).

Para Scriven es importante dedicar atención a la lógica de la evaluación para mejorar su práctica profesional. Para este autor, evaluar programas implica tres actividades cruciales:

1. determinar criterios, esto es, las dimensiones en las que el programa es evaluado, para lo que recomienda comprender bien su naturaleza: «entender lo que es un reloj permite automáticamente entender los criterios o dimensiones adecuados para juzgar su valor o mérito: exactitud, legibilidad, solidez...».
2. determinar los niveles de rendimiento aceptables en cada una de las dimensiones, utilizando estándares absolutos o comparativos;

3. medir el rendimiento para el caso concreto, comparando los niveles alcanzados con los de referencia. Para determinar si vale la pena hacer el análisis habrá que considerar «el tiempo y recursos necesarios para la investigación, la probabilidad de obtener alguna respuesta y la probabilidad de que la respuesta sea correcta» (1980: 45).

Todo ello lleva a elaborar un juicio de valor sobre el objeto evaluado. Para Scriven, la cuestión fundamental es determinar si se trata de un buen programa, si vale la pena mantenerlo, qué aspectos funcionan adecuadamente y qué es lo que resuelven. La mejor manera de obtener respuestas a cuestiones sobre efectos la proporcionan los métodos experimentales. Sin embargo, no es un gran defensor de estos en la medida que, a menudo, la adherencia al método lleva a responder preguntas irrelevantes. En estos casos prefiere una respuesta menos refinada, pero a una cuestión más relevante.

2. Campbell. La sociedad experimental

La obra de Campbell tiene un ámbito de aplicación que va más allá de la evaluación de programas. Sin embargo, su trabajo junto con Stanley, *Experimental and Quasi-Experimental Designs for Research* (1963) ha sido considerado como el que mayor influencia ha tenido en el desarrollo del campo (Shadish y Epstein, 1987). En él, Campbell y Stanley presentan una teoría de la causación, que se estructura sobre la base de los tipos de validez y de las posibles amenazas a la validez de una investigación y definen los diseños cuasiexperimentales como un tipo de investigación más factible al poder llevarse a cabo en contextos reales y no depender de la normalmente difícil asignación aleatoria de los tratamientos. De hecho, Campbell y Stanley analizan las funciones que cumple la asignación aleatoria en los verdaderos experimentos para explorar cómo se pueden conseguir de otra manera.

La teoría de la causación de Campbell se diferencia de las llamadas teorías esencialistas en que no pretende predecir de forma absoluta una variable dependiente (los resultados de un programa), sino simplemente estimar el impacto marginal de una variable independiente (programa) en aquella, cuando otras posibles causas se mantienen constantes a través del proceso

de aleatorización. Campbell y Stanley adoptan el término *validez interna* para referirse a las inferencias que puedan relacionarse con agentes causales manipulables y explican nueve amenazas a la validez interna, dedicando una especial atención a la selección y a sus subcategorías —maduración e historia—.

Su principal contribución fue, sin embargo, la descripción y evaluación de algunos diseños sin aleatorización —cuasiexperimentos—, pero que, sin embargo, facilitan la inferencia causal. Estos se fundamentan en dos estrategias: primero, la utilización de medidas pretest (cuanto más larga sea la serie de observaciones pretest, mejor) y segundo, la utilización de grupos de comparación lo más similares que se puedan formar como base de referencia para el no tratamiento. De la combinación de estas dos estrategias resultan múltiples posibilidades de diseños de investigación aplicables en la práctica.

Campbell prioriza la validez interna por encima de la validez externa. ¿De qué nos sirve generalizar una relación si dudamos de la existencia de esta? Pero Campbell es quien adopta el término *validez externa* y explica siete amenazas que pueden circunscribir una relación de causalidad a las características de un grupo, a una localización, momento o forma de desarrollar un programa. Campbell y Stanley advierten, sin embargo, sobre las limitaciones de los cuasiexperimentos y dejan bien claro que la mejor opción es para ellos la del experimento aleatorio.

Una buena parte del trabajo posterior de Campbell con otros investigadores y singularmente con Cook (1979) consiste en el análisis de la aplicación de los cuasiexperimentos. En un trabajo con Boruch (1975), advierte sobre las limitaciones del diseño cuasiexperimental más utilizado —el diseño de pretest-postest con grupos de comparación no equivalentes— y presenta una lista de experimentos desarrollados en ciencias sociales con el objeto de demostrar su factibilidad. Campbell defiende los experimentos argumentando que las dificultades que presentan en términos de costo, complejidad técnica y desarrollo temporal lento son superables.

Para Campbell, la administración debería evaluar una nueva política sometiéndola a un experimento antes de hacerla extensiva a toda la

población. Ello permitiría someter a prueba los programas claramente innovadores —dada la escasa relevancia política de una prueba piloto—, además de permitir establecer inferencias causales más claras.

El optimismo inicial de Campbell disminuiría a raíz de algunas críticas al método experimental en el contexto político-administrativo y ante la falta de voluntad política para llevar a cabo verdaderas reformas en el campo social y dar contenido, más allá de la retórica política, a la obligación de evaluar las políticas impulsadas por la legislación federal (1982: 118). En *Reforms as Experiments* (1969), defiende aquellos cuasiexperimentos que mejor pueden informar sobre relaciones causales. Concretamente, destaca el diseño de series temporales interrumpidas donde la intervención pública puede cambiar la tendencia de la serie y el diseño de discontinuidad en la regresión donde se asigna un determinado tratamiento sobre la base de puntos de corte en una variable cuantitativa.

Reforms as Experiments fue criticado por Weiss y Rein (1970) quienes encontraron dificultades en la implementación de la evaluación experimental de un proyecto piloto. Sobre la base de observaciones cualitativas, Weiss y Rein critican las evaluaciones que no informan sobre lo que ocurre en la «caja negra», refiriéndose con este término al proceso de implementación del programa. Campbell (1970), sin embargo, tiene una posición postpositivista y rechaza la idea de una teoría neutral o de un conocimiento objetivo. Al contrario, defiende que la medida cuantitativa se basa en suposiciones cualitativas: distintos actores prefieren distintos resultados y ello tiene consecuencias diversas en términos de conceptualización teórica. Por ello, rechaza la investigación que se fundamenta en una sola medida de resultados y, al contrario, propone la utilización de múltiples indicadores y el análisis de los procesos para superar el problema de la caja negra. Campbell no discute pues la oportunidad de los análisis cualitativos, pero sí se pronuncia en contra de que estos substituyan a los métodos cuantitativos en la función de análisis causal.

Después de *Reforms as Experiments*, Campbell escribió sobre la llamada «Sociedad Experimental», un artículo cuya publicación definitiva retrasaría unos años, en el que defiende su visión de una sociedad

racional en la que las decisiones se toman después de evaluar distintas alternativas en la forma de hacer frente a los problemas sociales. Campbell cree que la formulación de las alternativas depende de los grupos políticos. Es el proceso político el que ha de asignar pesos a los indicadores y sacar conclusiones. El papel del científico social no debe ser pues el de consejero sino el de metodólogo. *The Experimenting Society* (1988), que es como se acabaría titulando el artículo, se fundamenta en los principios de la multiplicidad de análisis, crítica abierta y réplica de los estudios siempre que sea posible.

3. Weiss. El contexto político y organizativo. La utilización de los resultados de la investigación social

Weiss es una socióloga que adopta las teorías y métodos experimentales en su primera obra, pero que pronto advierte sobre las dificultades políticas y organizativas que presenta la evaluación para producir resultados que sean efectivamente utilizados. Esta aproximación es desarrollada, entre otros, por Chelimsky (1987) y Palumbo (1987).

Weiss describe el pesimismo y la decepción que producen los resultados de la primera oleada de evaluaciones en los años 70. «Nada parecía funcionar como se esperaba (...) Ocasionalmente se detectaban efectos positivos. Algunos programas mostraban pequeños beneficios (...) Pero incluso las evaluaciones más recientes, utilizando criterios racionales y métodos de estudio más sofisticados tienden a encontrar éxitos marginales» (1987: 41). Por otra parte, «los datos recogidos parecían no tener ningún efecto en las decisiones presupuestarias o sobre la expansión o reducción de los programas» (1987: 42).

La teoría de Weiss responde a esta preocupación tomando como hilo conductor el problema de la insuficiente utilización de los resultados de la evaluación y de la investigación social en el proceso de formulación de las políticas. «En su forma ideal, la evaluación se lleva a cabo para un cliente que tiene que tomar decisiones y que mira a la evaluación en busca de respuestas en las que fundamentar sus decisiones» (1972b: 6).

Para Weiss, existe una incompatibilidad básica entre investigación social y programación social que explica los pobres resultados en términos de utilización. Su argumentación se fundamenta en las siguientes ideas:

1. La investigación no es el primer objetivo de los programas sociales. Cuando las necesidades de la evaluación y las necesidades del programa entran en conflicto, la prioridad es para el programa.
2. Los objetivos de los programas tienden a ser generales, difusos, diversos e inconsistentes. Varían para cada una de las partes afectadas y tienen poco que ver con el funcionamiento del programa. «A menudo, es preciso obtener el apoyo de una coalición para garantizar la aprobación de un programa. Hay que ganar la confianza de actores con valores e intereses distintos lo que lleva a hacer promesas realistas y no tan realistas» (Weiss, 1973a: 181).
3. Los técnicos y el personal de un programa suelen mostrar resistencia ante una evaluación. Su principal objetivo es producir el servicio mientras que una evaluación puede ser un obstáculo a sus tareas. Pueden pensar que la evaluación no es necesaria o que se les está juzgando a ellos, con lo que aún colaborarán menos.
4. Los grupos de control son factibles en laboratorios, pero son más difíciles de organizar en un contexto real. La administración puede rechazar la asignación aleatoria de tratamientos por razones éticas o de otro tipo, puede que todas las personas que precisen el tratamiento lo reciban y, aun en el caso de que no haya suficientes recursos para todos, cuando se produce una baja, la administración tiene que aceptar a una persona que estaba en el grupo de control.
5. En un laboratorio se puede homogeneizar el tratamiento que recibe cada individuo. En un contexto real, una prestación puede variar significativamente de un caso a otro, de un lugar a otro, «incluso en el caso de que el impacto medio no sea significativo, algunos lugares, algunas estrategias o determinadas formas de operar pueden funcionar» (1973a: 182).
6. Muchos programas cambian de forma y de contenido a lo largo de su existencia, con nuevos directores, nuevas reglamentaciones o cambios en la ideología dominante. Algunos programas son simplemente

un «conglomerado de actividades cambiantes que requieren un gran esfuerzo de especificar y describir» (1972a: 9). Si resulta difícil describir un programa, aún lo es más explicar qué componentes son responsables de determinados efectos.

7. En un laboratorio se puede esperar a que un tratamiento acabe o, por el contrario, si algo va mal, se puede cambiar. Un programa real raramente se puede cambiar completamente. Las innovaciones deben ser consistentes con la ideología y con los intereses del contexto político-administrativo. Por ello, los cambios serán «incrementales antes que masivos» (1980: 25).

Weiss advierte asimismo sobre el tiempo que a menudo debe transcurrir para que se puedan detectar los efectos de una intervención sobre un problema social. Por ello, estima poco realistas algunas evaluaciones que esperaban detectar resultados de forma inmediata. Aun otras veces, se suponía que el programa tendría un impacto tan fuerte que cambiaría la vida de las personas por encima de la influencia de otros factores.

Todo ello justificaría un cierto abandono de los métodos experimentales así como la adopción de métodos cualitativos. Sin embargo, estos también presentan sus limitaciones. El principal inconveniente reside, para Weiss, en la falta de credibilidad de estas aproximaciones. Sin evidencia empírica dura, los resultados de una evaluación serán interpretados como las «opiniones de un observador» (1987: 43). Weiss recomienda utilizar métodos cuantitativos y cualitativos. Estos últimos pueden aportar la flexibilidad, dinamismo y variedad de perspectivas de la que carecen los primeros.

La teoría sobre la utilización de los resultados de las evaluaciones presta una gran atención al contexto político de los programas sociales. «La evaluación es una empresa racional que se desarrolla en un contexto político» (1973b: 37). Las consideraciones políticas tienen entrada de tres formas distintas:

1. Los programas son el resultado de decisiones políticas (...) Son objeto de las presiones, a favor y en contra, que surgen a raíz del proceso político.

2. Los resultados de una evaluación entran en la arena política en la que compiten por la atención de los decisores con otros factores que tienen su peso en el proceso político.
3. La evaluación en sí misma es política en tanto que implícitamente adopta determinadas posiciones políticas sobre la legitimidad de los objetivos o de la estrategia del programa, el carácter discutible de algunas cuestiones y no de otras (...) «Parece que la investigación en evaluación tiene más posibilidades de afectar decisiones cuando el investigador acepta los valores, suposiciones y objetivos del decisor» (p. 41).

Otro elemento clave del contexto en el que se desarrolla la evaluación de programas es la existencia de múltiples partes afectadas por un programa. Un *stakeholder* es definido como el miembro de un grupo que es afectado de forma clara por el programa —y que por tanto puede verse afectado por las conclusiones de una evaluación— o el miembro de un grupo que toma decisiones sobre el futuro de un programa (1983a: 84).

Cada grupo tiene intereses distintos en el programa, se hace preguntas distintas sobre el programa y tiene distintos niveles de poder para utilizar los resultados de una evaluación. Muchas partes no tienen ningún interés en descubrir si el programa funciona. Esta diversidad de *stakeholders* y de intereses hace más difícil la utilización de los resultados de las evaluaciones.

«Algunos justifican el programa porque "hace algo" en un área en la que las necesidades son obvias. Otros se declaran satisfechos si el programa evita que un joven cometa un delito o prepara a un alumno para ir a la universidad. La administración, el personal, la clientela que se forma en torno a un programa tienen un interés en su perpetuación y en la expansión de la intervención» (1972a: 4).

Ante esta visión más bien pesimista de las posibilidades de utilización de las evaluaciones, Weiss orienta su trabajo, que desarrolla en parte con Bucuvalas, a analizar cómo se formulan las políticas y qué papel juega en este proceso la investigación social en un sentido más amplio. Para Weiss y Bucuvalas, raramente se utilizan los resultados de un estudio para dar respuesta a una pregunta específica que luego se implementa.

Más bien se produce una absorción de conceptos y generalizaciones de muchos estudios a lo largo de períodos de tiempo largos de manera que se produce una integración de los resultados de las investigaciones, junto con otras informaciones, en la interpretación que hacen los líderes políticos y administrativos de los hechos reales. Se produce así una «sensibilización gradual a las perspectivas de la investigación social» (1980: 263).

Weiss defiende frente a una utilización instrumental de la investigación una utilización conceptual. Su *enlightening model* parte de la idea que la investigación no resuelve problemas, sino que proporciona evidencias que pueden ser utilizadas para encontrar soluciones. En cada caso concreto, sin embargo, el investigador deberá considerar las llamadas tres I (de ideología, intereses presentes e información existente), ver cómo interaccionan y determinar si tiene sentido realizar un nuevo estudio (1983b).

Finalmente, Weiss distingue tres etapas en toda investigación sobre un programa:

1. Formulación de las cuestiones que deben ser investigadas. Su recomendación es la de leerlo todo, hablar con todo el personal, observar el programa y colaborar con todas las partes con el fin de alcanzar un acuerdo. En esta etapa, el investigador debe construir un modelo de los procesos que el programa pretende generar.
2. Realización del estudio. Weiss señala aquí la importancia de la calidad de los estudios para la posterior utilización de sus resultados. En su primera época hablaba más de experimentos y cuasiexperimentos. Posteriormente, acepta cualquier método (sondeos, estudios de caso, observaciones...), mientras se utilicen adecuada y correctamente. Desde la perspectiva de la organización de la investigación, el elemento crítico es, para Weiss, obtener la colaboración de las personas que trabajan en el programa, recogiendo las recomendaciones que se encuentran en la literatura.
3. Discusión de las implicaciones de la investigación para las políticas analizadas, señalando aquí la importancia de las recomendaciones y de la difusión de los resultados de la investigación (1972b).

4. Wholey. Evaluación para la gestión

La mayor parte de la obra de Wholey se refiere a la evaluación como una respuesta a la necesidad de los directivos y gestores públicos de evaluar sus programas con el objetivo de encontrar maneras de gestionarlos mejor. Abramson y Bellavita (en Wholey, Abramson y Bellavita, 1986) y en general, la literatura sobre *performance management* y control de gestión (Brizius y M.D. Campbell, 1991; Dilulio, 1994) siguen esta aproximación.

La prioridad para Wholey (1983) está en la gestión de los programas. La evaluación debería servir para facilitar mejoras evidentes en la gestión, rendimiento y resultados de los programas de la administración. El destinatario son los directivos públicos, en un sentido amplio, que incluye a gestores y a otros directivos como los responsables presupuestarios o los mismos legisladores.

Al igual que Weiss, Wholey justifica la evaluación en términos de su utilidad: «El esfuerzo por evaluar un programa solo se justifica si resulta en productos que se utilizan» (Abramson y Wholey, 1981: 42). Sin embargo, Horst, Nay, Scanlon y Wholey (1974) encontraron poca evidencia de que la evaluación de programas sirviera para mejorar los programas de la administración federal. Para Wholey, los problemas detectados en la gestión y evaluación de los programas son inherentes a la complejidad del contexto administrativo y a las diferencias en la implementación de un lugar a otro. «La evaluación de grandes programas es difícil por el costo de recoger datos y por la variación en las actividades y objetivos a nivel local» (Wholey, 1983: 124).

A ello hay que añadir el hecho de que la mayor parte de los programas son administrados por administraciones estatales o locales. «Muchos programas federales son grandes sobres de dinero para invertir en una área de problema (...) Un programa puede tener suficiente entidad para ser descrito por la prensa, para generar presión a favor o en contra de su existencia, o para ser adoptado por la administración federal y, sin embargo, las acciones que engloba esta intervención pueden no estar suficientemente especificadas» (Horst, *et al.*, 1974: 303).

El principal problema para Wholey es de gestión. «Si definimos gestión como la aplicación de recursos a tareas con el objeto de conseguir objetivos específicos», no se puede gestionar sin objetivos claros, hipótesis plausibles y demostrables sobre los resultados que se persiguen y sin la motivación, habilidad o autoridad para gestionar. La actividad de evaluación está estrechamente relacionada con la gestión en el modelo de gestión orientada a resultados de Wholey. Concretamente, las actividades de evaluación se desarrollan en seis niveles:

1. En el trabajo con los niveles político y administrativo y con otros grupos de interés para identificar objetivos y prioridades y diagnosticar las potencialidades de un programa. En este nivel se llegan a acuerdos sobre los objetivos del programa y sobre los indicadores que serán utilizados para valorar el rendimiento del programa y sus resultados.
2. En el desarrollo de los sistemas necesarios para analizar el rendimiento del programa en función de los objetivos e indicadores de rendimiento especificados.
3. En el trabajo con decisores y otros grupos de presión para fijar un nivel realista de mejora posible del rendimiento del programa.
4. En el desarrollo de sistemas que faciliten la utilización de la información generada con el objeto de incentivar y premiar dentro de la organización, las mejoras en el rendimiento y en los resultados del programa.
5. En la valoración de los resultados observados y en su documentación, señalando cómo se consiguen unos buenos resultados, así como los factores que pueden ser un obstáculo a la consecución de mejores resultados.
6. En la comunicación del rendimiento del programa a los niveles políticos y a la opinión pública en general (1983: 202).

No todos los programas requieren, sin embargo, todas estas actividades de evaluación. Un estudio de evaluación solamente debería encargarse cuando «la utilidad probable de la nueva información supera los costos de obtenerla» (1983: 119). Wholey propone cuatro tipos de estudio que pueden utilizar los gestores para determinar progresivamente si precisan una evaluación más intensa, amplia y compleja.

a. Análisis de evaluabilidad (Evaluability assessment)

El objetivo es responder a la cuestión de si una evaluación contribuirá a mejorar el rendimiento de un programa (1983: 35). Los productos del análisis de evaluabilidad son dos: un modelo del programa que asocie recursos, actividades, resultados y relaciones causales, y un conjunto de indicadores de rendimiento sobre los que hay un mínimo acuerdo (1983: 42).

Además de clarificar la «teoría» del programa, el análisis de evaluabilidad sirve para explorar la realidad del programa y para asistir a directivos y legisladores a determinar opciones de cambio y de mejora.

b. Evaluación rápida (Rapid feedback evaluation)

Un paso intermedio entre el análisis de evaluabilidad y la evaluación intensiva es la evaluación rápida con un doble objetivo: primero, hacer una valoración rápida del rendimiento del programa en términos de los objetivos e indicadores de rendimiento acordados; y segundo, proponer algunos diseños para una evaluación más global, válida y fiable (1983: 119).

La principal idea aquí es reducir el tiempo de *feedback* para los directivos y responsables políticos. La evaluación rápida solamente se puede hacer sobre la base de datos existentes, investigaciones previas, auditorías o informaciones recogidas en entrevistas, sondeos o visitas.

Otros autores (Hendricks, 1981) desarrollan los llamados *Service delivery assessments*, con la idea de producir información sobre un servicio sin las limitaciones que impone desarrollar un paquete de objetivos e indicadores previamente. Se trata de la descripción de las operaciones desarrolladas a nivel local, de su comparación con niveles de rendimiento considerados normales, de recoger información sobre problemas detectados y sobre las «mejores prácticas observadas», para hacer recomendaciones sobre posibles mejoras operativas (1983: 144).

c. Control de resultados (Performance monitoring)

Se corresponde con el nivel 2 de gestión orientada a resultados. Consiste fundamentalmente en el establecimiento de un sistema de control de resultados sobre la base de un acuerdo a nivel político y administrativo de los objetivos e indicadores que se consideran realistas y aceptables (1983: 155).

Wholey puntualiza que se trata simplemente de fijar la idea de rendimiento o resultados y medir el avance hacia los objetivos acordados, dejando para otros la tarea de estimar la medida en que las actividades del programa han causado los resultados estimados (1983: 155). Wholey entiende que un sistema de control de resultados es útil aun sin la posibilidad de establecer inferencias causales válidas. Se trata de un «sistema de señales» que sirve para el ajuste y la mejora continua del programa. Por otra parte, Wholey estima que, junto con los estudios de evaluación cualitativos, el control de resultados suele ser la alternativa más factible (1983: 155).

Con un sistema de estas características existe la posibilidad de comparar datos con un período previo, con el rendimiento esperado si se han desarrollado estándares o entre distintas unidades de un mismo programa. Wholey advierte sobre la posible «corrupción» de los datos si se utilizan para tomar decisiones que pueden afectar negativamente a los que los proporcionan. También advierte sobre la necesidad de llegar a acuerdos sobre las actividades que serán objeto de control.

d. Evaluación intensiva

La evaluación intensiva es el estudio de evaluación más complejo posible. La principal preocupación es aquí la del diseño de la investigación (experimental o cuasiexperimental) con el objeto de determinar la validez de las hipótesis causales que relacionan las actividades de un programa con sus resultados (1979: 49).

Sin embargo, es claro que para Wholey el control de resultados a través de un sistema de indicadores es la opción más factible, menos cara y más útil para el gestor (1983: 117). En el mismo sentido, una evaluación rápida

puede ser la mejor opción antes de entrar en un proyecto de evaluación intensiva que puede no merecer el esfuerzo en tiempo y dinero.

Respecto a la organización de la evaluación, Wholey señala la oportunidad de integrar la oficina de evaluación con el programa, así como la importancia de aportar resultados pronto, mantener contactos frecuentes y, en general, compartir de forma constante, datos, resultados y conclusiones con el personal y los directivos del programa.

5. Stake. Estudios de caso y métodos cualitativos

Stake es uno de los primeros defensores de los métodos cualitativos en evaluación (1975a, 1975b, Stake y Gjerde 1974) y uno de los autores que más ha hecho por su legitimación. Guba y Lincoln (1981, 1989), Lincoln y Guba (1985), House, (1980), Patton (1980) siguen esta aproximación, desarrollando una de las líneas de investigación más seguidas en la práctica profesional. Al igual que Weiss y Wholey, Stake reacciona ante las primeras señales de no utilización de los resultados en evaluación. Su línea de argumentación pasa por defender la metodología de los estudios de caso como instrumento para mejorar la práctica de la evaluación a nivel local. Stake rechaza que las ciencias sociales puedan descubrir relaciones de causalidad *input-output* y no presta atención a los indicadores. Para Stake, los científicos sociales deben describir situaciones individuales que puedan ser comprendidas por los no científicos.

Al igual que Scriven, Stake deja que la evaluación emerja de la observación del programa. «Desgraciadamente, solo algunas cuestiones pueden ser estudiadas en una evaluación formal. El evaluador debe decidir en qué centra su atención; para ello, ¿debe basarse en sus percepciones?, ¿en los objetivos formales del programa?, ¿en las actividades del programa?, ¿en las reacciones de los participantes en el programa? La mayoría se basarán en una noción preconcebida de éxito. Mi recomendación es prestar atención a lo que ocurre en el programa y después decidir las cuestiones de valor y los criterios» (Stake, 1975b: 15).

A diferencia de Wholey, Stake no orienta la evaluación en función de los intereses de los directivos encargados de la gestión. Su idea fundamental

es la de identificar los intereses que distintas personas o grupos tienen en el programa y servir a los que el programa debería ayudar. También tiene una preocupación fundamental por ayudar al profesional que trabaja día tras día en el programa. El trabajo del evaluador no es pues el de hacer un juicio sumativo. «Veo al evaluador en el papel de prestador de un servicio antes que en el rol de un filósofo-educador» (1975a: 36).

Esta concepción de la evaluación como un servicio está en la base de su propuesta de un modelo de evaluación que responda a las necesidades de los clientes. Por *responsive evaluation*, entiende aquella aproximación que sacrifica precisión en la medida para aumentar la utilidad de los resultados de la evaluación para los directamente implicados en el programa (1980a). Para Stake, la evaluación debe centrarse en las actividades del programa —no tanto en las intenciones—, responder a las necesidades de información de las partes implicadas y hacer referencia a las distintas perspectivas sobre los valores presentes al informar sobre el éxito o el fallo del programa (1980a).

Este modelo presenta las siguientes ventajas sobre la evaluación cuantitativa clásica:

1. permite que emerjan las variables importantes para el programa. Tradicionalmente se requiere al evaluador que especifique qué tratamientos y variables son más importantes. Stake entiende que ello no puede hacerse sin haber observado el programa o sin un cierto grado de flexibilidad para adaptar cambios en el programa.
2. facilita los cambios en el programa. Para Stake, las evaluaciones desarrolladas en relación con la reforma educativa en los Estados Unidos han tenido un efecto desilusionador cuando la evaluación debería impulsar los cambios y las reformas. «El evaluador hace su mejor contribución cuando ayuda a descubrir ideas, respuestas, soluciones que muchas veces están en la mente de los que participan de una forma u otra en el programa» (Stake, 1975a: 36).
3. devuelve el control al nivel local. Mientras que el modelo de evaluación clásico está dando el mensaje contrario «eres incompetente, necesitas ser dirigido» (1980b: 161), Stake anima a los evaluadores a resistirse a aceptar los objetivos y criterios de las administraciones estatal y

federal y a evitar que la evaluación sirva para dar mayor poder a determinados *stakeholders* (1979: 56).

Las recomendaciones a nivel metodológico pueden resumirse en las siguientes ideas:

1. dedicar mucho tiempo a la observación del programa, reunir valoraciones, ver qué necesidades tienen los clientes y preparar comunicaciones de carácter informal;
2. dejar que las cuestiones emerjan y cambien durante la evaluación. El contacto intensivo con el programa permite descubrir las cuestiones que van a estructurar las discusiones con los clientes;
3. cuidar la comunicación con las personas interesadas en la evaluación. Los resultados deberían expresarse de una forma natural que permita a su audiencia asimilar la información y comprender su significado (1980a: 83).

Stake muestra una clara preferencia por los estudios de caso. «El caso no tiene que ser una persona o un proyecto. Puede ser cualquier sistema con unos límites claros» (1978: 7). Para Stake, un observador es el mejor instrumento para valorar muchas cuestiones que aparecen en una evaluación al poder ejercer su juicio crítico con mayor capacidad de apreciación de la realidad y de los resultados de un programa. La observación directa del caso es pues el elemento crucial de la evaluación. Para ello, los observadores pueden utilizar entrevistas, observación, examen de archivos y documentos, entre otros instrumentos metodológicos (1980a).

Para Stake, el objetivo de la evaluación es mejorar la práctica diaria del programa objeto de estudio y no producir nuevos conocimientos formales que puedan ser difundidos. En este contexto, el estudio de caso es relevante, dado que puede generar un proceso de aprendizaje que se basa en el conocimiento tácito y en la asociación de ideas permitiendo «comprender las cosas de una forma distinta, ver nuevos significados y nuevas aplicaciones de lo antiguo» (Stake, 1978: 6). La base para la llamada «generalización natural» la pone el lector del caso al reconocer aspectos esencialmente similares a los de su experiencia (1978: 7). El lector

tiene la capacidad de reconocer lo que es relevante e irrelevante para sus circunstancias particulares. Stake reserva pues un papel importante a la experiencia y a las intuiciones de los profesionales interesados. La responsabilidad del investigador es proporcionar los datos contextuales que faciliten este poder intuitivo del lector. Por otra parte, el estudio de caso proporciona información que puede ser interpretada de distintas maneras sin favorecer ningún punto de vista por encima de los demás.

Los estudios de caso son especialmente adecuados para el modelo de evaluación de Stake. Sin embargo, presentan algunas dificultades que el mismo Stake comenta:

El problema de *la validez de un estudio de caso*. La comprobación de la validez se puede conseguir con la triangulación de distintas investigaciones, intentando llegar a un mismo resultado por tres vías independientes. Esta idea es aplicable a un mismo estudio de caso o a distintos estudios de caso para un mismo programa en distintos puntos geográficos. Por otra parte, Stake entiende que el método de la «generalización natural» tiene la ventaja de que los lectores de un caso pueden participar en la determinación de su validez (Stake y Easley, 1978: 27).

El problema de *la especificidad de los estudios de caso*. Stake entiende que los estudios de caso no son la alternativa más adecuada cuando se requiere un cierto nivel de abstracción o se pretende formular leyes generales y desarrollar teorías científicas. Aun así, reivindica la utilidad de los estudios de caso para la función de exploración que toda actividad científica requiere (1978: 7).

6. Cronbach. Innovación metodológica y capacidad de generalización

Cronbach escribe sobre evaluación de programas después de haber dedicado una parte importante de su vida a temas metodológicos. Son importantes sus trabajos sobre fiabilidad (1951), validez interna (1955) y generalización (Cronbach y Snow, 1977). A mitad de los años 70, su visión de la teoría y métodos en ciencias sociales había evolucionado considerablemente. Adopta el término social inquiry, en lugar de ciencias sociales, para poner de relieve la necesidad de métodos interpretativos

y descriptivos que puedan complementar los métodos tradicionales de muestreo, experimentación y análisis cuantitativo. Basándose en su propia experiencia, Cronbach ve la necesidad de descubrir nuevos métodos que se adapten a la naturaleza compleja y multivariable del contexto social.

En *Toward Reform of Program Evaluation* (1980), Cronbach y sus colaboradores de la Universidad de Stanford utilizan investigación empírica y teoría del análisis de políticas y de la evaluación de programas para presentar una teoría de la evaluación que pone el énfasis en la utilización de los resultados a corto plazo, tiene en cuenta el contexto político y administrativo caracterizado por la pluralidad de *stakeholders* y persigue satisfacer una necesidad de conocimiento que pueda ser extrapolado a otras situaciones.

Cronbach y otros (1980) entienden que buena parte de los problemas de uso de las evaluaciones provienen de modelo racional en que se fundamentan. Para estos autores, normalmente, a) no se toman decisiones de un forma clara en un momento determinado; b) hay más de un decisor; c) la política es más importante que la información sobre cómo optimizar las decisiones; d) no se toman decisiones del tipo sí/no sobre la continuidad de un programa; e) no hay tiempo para prestar atención a todas las fases de una investigación; f) los informes llegan tarde; g) los decisores que piden un estudio y los que reciben sus resultados son distintos; h) los informes no están pensados para un gestor; i) los resultados de un estudio no se utilizan instrumentalmente para modificar un programa (...) En cambio, las políticas sociales son adoptadas cuando tienen un apoyo político suficiente, hasta el punto de poder tener una influencia relevante en votos. Los objetivos tienen que ser lo bastante vagos para poder acomodar múltiples intereses.

Todo ello sugiere la necesidad de incorporar el contexto político y administrativo a la teoría de la evaluación. Este entorno puede variar de forma considerable si se examinan las posibilidades que resultan de combinar actores y etapas de desarrollo de un programa. Concretamente, Cronbach y otros hablan de cinco grupos que influyen en el proceso de dar forma a una política (*policy shaping*): los responsables a nivel de formulación de la política (en el ejecutivo o en el legislativo), los responsables a nivel de programa («los que tienen el nombre del programa en su puerta»),

los prestadores de servicios a nivel local, los clientes y beneficiarios del programa, y los académicos, periodistas y otros grupos que interpretan o difunden información sobre el mismo. A su vez, distinguen entre tres etapas: una etapa inicial de probar nuevas ideas, una etapa de demostración en la que se maximizan las posibilidades de éxito y una tercera etapa en la que el programa funciona en condiciones normales.

La idea de *policy shaping* se basa en la teoría del cambio social gradual. De acuerdo con esta concepción, Cronbach y otros rechazan que la evaluación sirva directamente para tomar decisiones o para controlar el rendimiento de un gestor. La evaluación sirve para «iluminar» las decisiones sobre distintas alternativas posibles, clarificando las cuestiones que interesan a las partes afectadas. La idea principal es entender el problema y el programa, el contexto político y, sobre todo, qué procesos contribuyen a los resultados esperados. Esto último es particularmente importante para que estos procesos sean luego extrapolables a otras situaciones. Por ello, la teoría de Cronbach otorga un papel central a la validez externa.

Posteriormente, en *Designing Evaluations of Educational and Social Programs* (1982a), Cronbach ataca directamente los postulados de Campbell y sus colaboradores al defender un nuevo concepto de validez que tiene que ver con la capacidad de producir conocimientos que trasciendan una situación específica. Para Cronbach,

1. la validez no es una propiedad del diseño de una investigación, sino una propiedad de sus conclusiones, que tiene que ver con su credibilidad para las personas que podrían utilizar dichas conclusiones;
2. un tratamiento no se puede diferenciar de las unidades, observaciones y características contextuales en las que se presta y de las que su réplica depende;
3. existe una necesidad de extrapolar, esto es, de transferir los resultados de una investigación más allá del contexto en que se generan.

En cuanto al diseño de la evaluación, la propuesta de Cronbach se basa en tres principios:

1. no hacer un solo estudio de evaluación muy costoso, sino varios estudios pequeños que tengan una única racionalidad;
2. no dedicar muchos recursos a responder con precisión unas pocas cuestiones específicas, sino dar respuestas menos claras a una amplia gama de cuestiones;
3. priorizar en función de la relevancia de las cuestiones para la comunidad de actores asociada a la política y del grado de incertidumbre en el conocimiento sobre los temas propuestos.

El papel del evaluador en este contexto es el de un educador independiente, no partisano, que informa a las partes, transmitiendo el conocimiento existente y enseñando a pensar críticamente (1980: 67).

7. Rossi. Especificación de modelos e integración teórica

Rossi, un importante sociólogo que ha trabajado diversos aspectos de la metodología de las ciencias sociales, es una figura clave de la evaluación de programas tanto por sus trabajos de investigación aplicada (Berk, *et al.*, 1981; Rossi y Lyall, 1978), como por su trabajo conceptual de integración de las teorías desarrolladas por otros autores. Rossi es conocido por el libro de texto que escribió junto con Freeman y que ha sido uno de las principales referencias para la formación en evaluación de programas. Rossi presenta su aproximación integradora sobre la base de tres conceptos: la idea de evaluación total —*comprehensive evaluation*—, la evaluación a la medida —*tailored evaluation*— y la evaluación sobre modelo teórico —*theory-driven evaluation*—.

Frente a la dispersión sobre las cuestiones a las que una evaluación debería responder, Rossi formula la idea de una evaluación que agrupa tres tipos de actividades: el análisis de la conceptualización y del diseño de una intervención —es preciso documentar la necesidad de un programa para planificar, llevar a término y evaluar acciones sociales—; el control de la implementación de un programa —el desarrollo de un simple sistema de

indicadores para el control de actividades puede ser tan a más importante que producir información sobre su impacto—; y la evaluación de la utilidad del programa —salvo que los programas tengan un impacto demostrable, será difícil defender su continuidad—.

Dada la imposibilidad práctica de llevar a cabo una evaluación que desarrolle simultáneamente estas tres actividades, Rossi propone la idea de evaluación a la medida, con el objeto de adaptar cada evaluación al programa objeto de estudio. Rossi distingue entre programas nuevos, reformas o modificaciones de programas que ya funcionan y programas estables en pleno funcionamiento. Para cada estadio o nivel de desarrollo, será preciso determinar las cuestiones y los métodos más adecuados. Así, por ejemplo, en el caso de un programa nuevo, deberían dedicarse pocos recursos a estimar efectos. Únicamente tendría sentido dedicar esfuerzos a analizar los resultados de programas similares o a recoger las opiniones de expertos sobre los efectos que cabe esperar.

Rossi utiliza otro concepto, *theory-driven evaluation*, para reclamar mayor atención al desarrollo de modelos teóricos de las intervenciones sociales. En su trabajo con Chen, estos autores señalan la falta de atención en la literatura a la importante tarea de comprensión de las políticas sociales: «lo que es realmente atractivo de los experimentos controlados es que no es necesario entender cómo funciona un programa social para entender sus efectos netos (...)» (1983: 284). Pero si los resultados del experimento no muestran efectos, no se sabe si el fallo es debido a que el programa estaba construido sobre fundamentos conceptuales pobres o a que el tratamiento era claramente insuficiente para producir efectos o a que se han producido fallos en la implementación.

Rossi defiende la necesidad de especificar modelos sobre cómo cabe esperar que funcione un programa antes de evaluarlo. Si bien Rossi es claramente experimentalista en su primera época, posteriormente combina esta aproximación con la modelización estructural con el objeto de obtener información sobre los inputs, procesos intermedios y productos de los programas. Este esfuerzo por modelizar teóricamente la intervención social le permite integrar las aproximaciones dispares de Campbell y Cronbach a la cuestión de la validez. Para Rossi, se pueden tratar los problemas de validez

interna especificando un modelo que tenga en cuenta las relaciones entre variables de la forma más completa posible. La especificación de un modelo facilita asimismo la generalización en la medida que fuerza al investigador a definir el sistema en el que ha de operar el programa, las características contextuales y las posibles interacciones (Chen y Rossi, 1987).

En *Evaluation. A Systematic Approach*, Rossi y Freeman (1985) incorporan algunas consideraciones sobre la práctica de la evaluación.

Respecto a la cuestión sobre los *stakeholders* de un programa y la perspectiva que debería adoptar el evaluador, Rossi y Freeman otorgan una especial atención a los que están más implicados en la toma de decisiones. Sin embargo, el evaluador tiene que ser muy claro en lo que concierne a desde qué perspectiva se hace una evaluación y debe reconocer de forma explícita la existencia de otras perspectivas. Por otra parte, no debe incorporar los sesgos del que financia la evaluación, siendo esta una cuestión moral o de ética profesional.

Rossi prefiere las técnicas cuantitativas. Los métodos cualitativos suelen ser caros, difíciles de utilizar e imprecisos, según Rossi. Sin embargo, recomienda utilizar técnicas cualitativas para el análisis de la conceptualización, el diseño y para el control de las operaciones de un programa. En este caso, es partidario de estructurar de alguna forma la recogida de datos para poder presentar información de forma resumida y que permita su análisis.

Toda la discusión de la evaluación en *Evaluation. A Systematic Approach* se organiza sobre la base de la distinción entre programas innovadores, programas estables en funcionamiento y programas objeto de reformas y cambios. Al igual que Campbell, Rossi otorga mayor atención a los programas nuevos. La situación ideal para Rossi es la de trabajar con los responsables del programa optimizando las probabilidades de obtener una buena intervención y una buena evaluación. Respecto a este tipo de programas y evaluaciones, Rossi dedica una especial atención a las siguientes cuestiones:

1. La importancia de trabajar con objetivos específicos. Rossi discute distintas propuestas de otros autores, como los análisis de evaluabilidad de Wholey —en los que la idea principal es, sobre la base del acuerdo de los *stakeholders*, decidir criterios y niveles aproximados de éxito/fracaso. También recoge la posibilidad de incluir objetivos aceptados en la literatura del campo substantivo correspondiente.
2. Los modelos de impacto. Rossi distingue entre una hipótesis causal, una hipótesis de intervención y una hipótesis de acción. La primera se refiere a la relación causal, la segunda, a cómo la intervención pretende afectar esta relación y la tercera, a la posibilidad que la intervención no esté necesariamente relacionada con el resultado deseado, incluso en el caso que este se produzca.
3. El análisis de la forma de prestación del servicio. Para Rossi, es crucial examinar si el programa se ha llevado a la práctica correctamente. Esto supone analizar el problema y la población objetivo, los servicios prestados y sus componentes específicos, las características del personal, el proceso de selección de los beneficiarios y las formas de acceso al programa.
4. La posibilidad de experimentar. Lo que es más factible cuando se trata de programas nuevos, no extensivos a toda la población potencial, sino únicamente a determinados grupos de beneficiarios, lo que permite construir grupos de control y establecer comparaciones. Al igual que Campbell, Rossi prefiere los experimentos para la evaluación de impacto. Sin embargo, recomienda seleccionar el mejor diseño posible en función de las condiciones y circunstancias específicas de cada caso. «El evaluador debería escoger el mejor diseño posible, teniendo en cuenta que sea practicable y factible (...), los recursos disponibles y los conocimientos del evaluador» (1985: 167).

Respecto al problema de la utilización de las evaluaciones, Rossi defiende una utilización instrumental de sus resultados. Para que ello sea posible, cree necesario que los evaluadores entiendan la forma de trabajar de los decisores, lo que implica, primero, presentar los resultados a tiempo, cuando son necesarios, y de forma breve; segundo, ser sensible a las posiciones de todas las partes; y, tercero, planificar su difusión y utilización. También ve

la posibilidad de una utilización conceptual o de una utilización política, para atacar o defender determinadas posiciones (1985: 188).

III. El marco para la práctica de la evaluación

La evolución teórica de la evaluación es fruto de la práctica profesional durante los últimos veinticinco años. La teoría, a su vez, es útil como marco de referencia para una práctica profesional que a menudo precisa de orientación en su aproximación a los problemas concretos con los que se enfrenta. De la observación del cuadro 1, se podría concluir que hay una considerable diversidad de aproximaciones en la disciplina. Sin embargo, también se produce una cierta convergencia al irse aceptando algunos principios generales que pueden servir de guía para la práctica de la evaluación.

Así, un profesional que se enfrenta al reto de una evaluación tiene que optar entre múltiples alternativas, cuya argumentación teórica parece irrefutable. Pero, a la vez, tiene algunos criterios que puede seguir con una cierta seguridad cuando se plantea cuestiones críticas del tipo: ¿debería hacer esta evaluación?, ¿en qué debería centrarse?, ¿cómo se puede hacer?, ¿desde qué perspectiva?, ¿cómo se puede facilitar su utilización posterior?

Si la cuestión es qué evaluar, la teoría ofrece distintas posibilidades que van desde los efectos hasta las necesidades sociales, pasando por los procesos intermedios y los análisis costo-eficacia. La teoría ofrece asimismo distintos actores que pueden responder esta cuestión: directivos públicos, clientes, profesionales, las instituciones que financian la evaluación. Existe pues una variedad de respuestas posibles a la pregunta de quién decide o para quién se hace la evaluación.

El contrapunto a esta amplia gama de posibilidades está en algunos principios que indican al evaluador cómo se relacionan estas variables. Así, parece claro que cada actor tiene preocupaciones distintas en función del lugar, la responsabilidad o el interés que tiene en el programa. Sin embargo, independientemente de quien sea el cliente del estudio, el evaluador ha de tener en cuenta los valores u objetivos del conjunto de *stakeholders*, incluyendo a afectados directa o indirectamente por el programa y a

decisores con poder efectivo. Por otra parte, el evaluador deberá responder a las necesidades de información de los comanditarios de la evaluación si quiere facilitar su posterior utilización. A su vez, la opción sobre el objeto de la evaluación estará determinada por la etapa de desarrollo en la que se encuentra el programa, la naturaleza del programa, la información existente sobre este y los recursos —humanos y materiales— invertidos en la evaluación.

Si la pregunta que se plantea el evaluador es cómo hacer su estudio, las respuestas ilustran una división clara a nivel metodológico. La gama de posibilidades incluye métodos de tipo experimental, métodos cuantitativos clásicos (estadística descriptiva, análisis inferencial, correlación, regresión, sondeos), métodos de control de gestión (sobre la base de sistemas de indicadores), métodos cualitativos clásicos (observación, entrevistas, análisis de documentos) y estudios de caso.

Sin embargo, como se desprende de la teoría, la finalidad y el objeto de la evaluación determinan considerablemente las opciones metodológicas. Así, por ejemplo, existe un cierto acuerdo en que los estudios de caso no son la mejor opción para el análisis de efectos y de inferencias causales, o en que cuestiones de conceptualización y diseño de una intervención en un ámbito social sobre el que el conocimiento está poco estructurado requieren una aproximación más cualitativa.

Otra cuestión que puede tener respuestas plurales es la que se plantea la contribución que puede realmente esperarse de una evaluación. Aquí, los dos extremos estarían representados por las siguientes posiciones. Por una parte, se puede entender que la evaluación únicamente tiene sentido en cuanto resulta útil de forma inmediata, a corto plazo, de ahí, su financiación. Por otra parte, la evaluación será realmente útil en cuanto contribuya a producir un cambio en la forma de concebir una determinada realidad social y las posibilidades de intervención sobre ésta desde las instituciones públicas.

Cuadro 1. Modelos teóricos en evaluación de programas

| | SCRIVEN | CAMPBELL | WEISS | WHOLEY | STAKE | CRONBACH | ROSSI |
|---|-------------------------|---|---|--|---|---|--|
| Nuevos términos | Ciencia de valoración | Sociedad Experimental | Enlightening Model | Análisis de Evaluabilidad | Responsive, Naturalistic Model | Análisis de UTOS (Unidades, Tratamientos, Observación, Contexto) | Theory-Driven Model |
| Autores con una visión similar | | Boruch, Cook, Riechert, Rossi | Cronbach, Chelmsky, Palumbo, Wholey | Abramson, Bellavita, Horst, Noy, Scanlon | Guba, Lincoln, Patton | | Berk, Chen, Freeman, Wright |
| Concepto | Juicio de valor | Producir información para la toma de decisiones | Clarificar cuestiones, «iluminar» a decisores y afectados | Producir información para la gestión | Producir información para los miembros del programa | Clarificar cuestiones, sobre: problema, programa y contexto | Producir información para la toma de decisiones |
| Interés prioritario | Interés general | Responsables del programa | Legislador y directivos | Legislador y directivos | Beneficiarios del programa, profesionales | Legislador y directivos | Responsables del programa |
| Rol del evaluador | Juez | Metodólogo | Educador | Agente de cambio | Prestador de un servicio | Educador, Difusor | Metodólogo |
| Objeto | Efectos | Efectos Rel. Causales | Efectos, procesos, contexto | Efectos, actividades, productos | Actividades | Efectos, procesos, contexto | concepto, implementación, eficacia |
| Fuente de criterios para determinar eficacia | Necesidad social | Objetivos | Objetivos, stakeholder, documentación | Stakeholder | clientes, técnicos | Stakeholder | stakeholder, literatura científica |
| Métodos | Lógica experimental | Experimento Cuasexperimento | Métodos cuantitativos, Métodos cualitativos | Análisis de evaluabilidad Sistemas de indicadores, Evaluación rápida | Métodos cualitativos, Estudios de caso | Métodos cuantitativos Métodos cualitativos | Experimento, Cuasexperimento Métodos cualitativos, excepto relaciones causales |
| Perspectiva sobre utilización | Instrumental | Instrumental | Conceptual Formulación de políticas | Instrumental | Instrumental | Conceptual Comprensión de la realidad social | Instrumental Conceptual y Política Formulación de teoría |
| Oportunidad de la evaluación | Criterio costo-eficacia | Voluntad de reforma superada la etapa de iniciación | En función de los tres I: intereses, información disponible e ideología | Criterio costo-eficacia | Útil para los miembros del programa | En función de la relevancia del estudio, y de la falta de información | Programas innovadores fase de prueba piloto |
| Organización de la evaluación | Operación externa | Operación externa | Implicar al personal, no impedir la prestación del servicio | Evaluación parte del sistema de gestión por resultados | Contacto permanente con el programa | Obtener la cooperación del programa | Adaptarse al tiempo y forma de trabajar de los gestores |

Nuevamente, existe una posibilidad de integración de estas dos posturas sobre la base de diferenciar dos realidades diversas: la evaluación en tanto que servicio a los profesionales o directivos de un programa concreto, y la evaluación en tanto que oportunidad para la reflexión sobre una determinada forma de tratar una realidad social problemática. En la primera, no hay una preocupación teórica fundamental que si está presente en la segunda, donde la información específica sobre el detalle del funcionamiento del programa puede ser menos relevante. Normalmente, cabría esperar que los evaluadores académicos tengan un mayor interés por intervenir en las evaluaciones del segundo tipo, mientras que los consultores profesionales lo harán en las del primer tipo.

Otra pregunta que puede hacerse el evaluador es cuando tiene sentido hacer una evaluación. Las respuestas posibles van desde las prescripciones técnicas en función de los métodos empleados a consideraciones sobre el contexto político (intereses, ideología, propósitos ocultos) y administrativo (urgencia de la decisión, percepción de la evaluación por el personal afectado, medida en que la evaluación puede afectar la prestación regular del servicio) o sobre los recursos que requiere.

De hecho, todas estas consideraciones son perfectamente compatibles. Una característica que se desprende del estudio de la teoría de la evaluación es que se pasa de una consideración del programa en un vacío social a la progresiva incorporación del contexto y de la preocupación por adaptar la evaluación a un sistema político plural y descentralizado. La práctica de la evaluación no puede por tanto obviar que la evaluación es política y que sus resultados son una información más en el proceso político que tiene que competir con informaciones que tienen su acceso por otras vías.

Aun otra cuestión que a menudo se plantean directivos y evaluadores es cómo organizar la evaluación. Las prescripciones van desde la organización de la evaluación como una operación externa a la integración de la evaluación en el equipo que desarrolla el programa, pasando por la lógica formación de equipos mixtos. Cada una de estas opciones tiene sus ventajas y sus inconvenientes, mientras que no siempre se disponen de recursos suficientes para montar el mejor dispositivo de evaluación.

Sin embargo, la teoría es clara respecto a esta cuestión. Raramente se puede hacer una evaluación desde un despacho, sin el contacto con la realidad del programa. Es necesario pues mantener el contacto, alimentar el interés en la evaluación e implicar a las personas claves en las principales decisiones que se toman durante el proceso evaluador para facilitar su posterior utilización.

IV. Conclusiones

El repaso de las teorías de siete autores claves en evaluación de programas, siguiendo el trabajo de Shadish, Cook y Leviton (1991), ilustra la evolución de la disciplina y su progresiva estructuración sobre la base de otras ciencias sociales y de los principios extraídos de la propia experiencia en evaluación. Crane (1988) utiliza la metáfora de una «zona de pastos común» para poner de relieve el carácter interdisciplinario de la evaluación de programas. Cronbach y sus colaboradores (1980) hablan de una «industria de servicios» donde importa menos la afiliación que la capacidad para resolver problemas, poniendo así de relieve su carácter práctico.

Los teóricos de la evaluación han desarrollado de forma intensiva los aspectos metodológicos, dada la necesidad de dar respuestas creíbles a los problemas planteados por la administración. También, han dedicado una parte importante de su trabajo a desarrollar aspectos de proceso, evolucionando desde una concepción de la evaluación como ciencia a una nueva concepción de la evaluación como arte, en el sentido de habilidad para construir una investigación de calidad que resulte útil en un determinado contexto político-administrativo. Todo ello sin olvidar que evaluar implica necesariamente la idea de juicio de valor y por tanto la adopción de unos determinados principios que luego son traducidos a criterios de valoración.

La evolución teórica de la evaluación de programas es fruto de la práctica profesional. Por ello, se ha producido una diversificación considerable. Se ha pasado del estudio exclusivo de los efectos a la preocupación por examinar la aplicación concreta de los programas y los procesos causales que median en la producción de los impactos deseados. La confianza exclusiva en aproximaciones de tipo cuantitativo-experimental ha dado

paso a la inclusión de los métodos cualitativos y de los estudios de caso. Los legisladores y directivos públicos fueron en una primera etapa la única fuente de las cuestiones y de los criterios de evaluación, mientras que progresivamente, otras partes interesadas han sido incluidas en esta consideración. La utilización efectiva de las evaluaciones, un tema inicialmente ignorado, ha pasado a ser una de las principales preocupaciones en la disciplina.

La diversidad en las aproximaciones no es contradictoria con un proceso de integración de la teoría, que aún queda por hacer, pero que es evidente cuando se analiza la literatura. La evaluación no es pues una disciplina a teórica ni la teoría de la evaluación es un conjunto de proposiciones de escasa utilidad para la práctica de la disciplina. Al contrario, la evaluación tiene en la teoría emergente un conjunto de proposiciones que pueden dar respuestas útiles a las cuestiones críticas que plantea la práctica de la evaluación y de las que depende que ésta sea realmente útil para la reforma de las intervenciones públicas y la formulación de nuevas políticas.

Notas

* Título original “Modelos teóricos para la práctica de evaluación de programas”, en Quim Brugue y Joan Subirats (comp.), *“Lecturas de gestión pública”*, Ministerio de las Administraciones Públicas de España, Madrid, 1999. Publicación autorizada al español ©1999, Ministerio de las Administraciones Públicas en España. Reproducción autorizada.

Publicado en Brugué Q y Subirats J. (eds.) (1996). *Lecturas de gestión pública*. Madrid: Instituto Nacional de Administración Pública, (pp. 321, 352). ISBN: 84-340-0907-2.

¹ Ante la dificultad de hacer una selección, he preferido seguir el criterio de Shadish, Cook y Leviton. La falta de autores europeos no debería sorprender al lector. Las aportaciones teóricas originales en evaluación se deben fundamentalmente a autores norteamericanos. Los trabajos europeos en evaluación son fundamentalmente obras de divulgación o de aplicación a casos concretos. En España, los trabajos de Alvira (1992) o los míos propios (Ballart 1992, 1994; Ballart y Riba, 1996) entrarían dentro de estas categorías. Algunos artículos recientes de revisión de la evaluación en otros países europeos (Pollit, 1992, para Inglaterra; Duran, Monnier, Smith, 1995, para Francia) confirman esta apreciación.

Referencias

- Abramson, M.A. y Wholey, J.S. (1981). Organization and Management of the Evaluation Function in a Multilevel Organization. In R.J. Wooldridge (Ed.), *Evaluation of Complex Systems* (31-48). San Francisco: Jossey-Bass.
- Alvira Martín, F. (1991). *Metodología de la Evaluación de Programas*. Madrid: CIS.
- Ballart, X. (1992). *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*. Madrid: Ministerio Administraciones Públicas.
- Ballart, X. (1994). Clientes de un servicio cultural: una experiencia de evaluación con focus groups, *Gestión y Análisis de Políticas Públicas*, N.º 1.
- Ballart, X. y Riba, C. (1996) «Impact of legislation requiring moped and motorbike riders to wear helmets», *Evaluation and Program Planning*, N.º 1.
- Berk, R.A., Lenihan, K.J., & Rossi, P.H. (1981). Crime and Poverty: Some Experimental Evidence from Ex-Offenders. In H.E. Freeman & M.A. Solomon (Eds.). *Evaluation Studies Review Annual* (Vol. 6. pp. 339-359). Beverly Hills, CA: Sage.
- Brizius, J.A., Campbell, M.D. (1991). *Getting results: A Guide for Government Accountability*, Washington, DC.: Council of Governors' Policy Advisors.
- Campbell, D.T. (1969). Reforms as Experiments. *American Psychologist*, 24, 409-429.
- (1970). Considering the Case Against Experimental Evaluations of Social Innovations. *Administrative Science Quarterly*, 15, 110-113.
- (1982). Experiments as Arguments. In E.R. House, S. Mathison, J.A. Pearsol, & H. Preskill (Eds.). *Evaluation Studies Review Annual* (Vol. 7, 117-127). Beverly Hills, CA: Sage.
- (1988). *Methodology and Epistemology for Social Science: Selected papers* (E.S. Overman Ed.) Chicago: University of Chicago Press.
- Campbell, D.T., & Boruch, R.F. (1975). Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects. En C.A. Bennett & A.A. Lumsdaine (Eds.). *Evaluation and Experiments: Some Critical issues in Assessing Social Programs* (pp. 195-296). Nueva York: Academic Press.
- Campbell, D.T., & Stanley, J.C. (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Chelimsky, E. (1987). The Politics of Program Evaluation. En D.S. Cordray, H.S. Bloom y R.J. Light. *Evaluation Practice in Review*. San Francisco: Jossey-Bass.
- Chen, H. & Rossi, P.H. (1983). Evaluating with Sense: The Theory-Driven Approach. *Evaluation Review*, 7, 283-302.

- (1987). The Theory-Driven Approach to Validity. *Evaluation and Program Planning*, 10, 95-103.
- Cook, T.D., D.T. Campbell (1979). *Quasiexperimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- Crane, J.A. (1988). Evaluation as Scientific Research. *Evaluation Review*, 12, 467.
- Cronbach, L.J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297-334.
- (1982a). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. (1980). *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Bass.
- Cronbach, L.J. & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cronbach, L.J. & Snow, R.E. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. Nueva York: Irvington.
- Dilulio Jr, J.J. (1994). Measuring Performance When There is No Bottom Line. *Center for Public Management Working Paper 94-2*, Washington DC: The Brookings Institution.
- Duran, P., Monnier, E., Smith, A. (1995). Evaluation a la française. *Evaluation N.º 1*.
- Guba, E.G. & Lincoln, Y.S. (1981). *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*. San Francisco: Jossey-Bass.
- Guba, E. G. & Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Beverly Hills, CA: Sage.
- Hendricks, M. (1981). Service Delivery Assessment: Qualitative evaluations at the cabinet level. In N.L. Smith (Ed.) *Federal Efforts to Develop New Evaluation Methods* (pp. 5-24). San Francisco: Jossey-Bass.
- Horst, P., Nay, J.N., Scanlon, J.W., & Wholey, J.S. (1974). Program Management and the Federal Evaluator. *Public Administration Review*, 34, 300-308.
- House, E.R. (1980). *Evaluating with Validity*. Beverly Hills, CA: Sage.
- Lincoln, Y.S., Guba, E.G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Palumbo, D.J. (1987) *The Politics of Program Evaluation*. Beverly Hills, CA: Sage.
- Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills, CA: Sage.
- Pollit, C. (1992). *Occasional Excursions: A Brief History of Policy Evaluation in the UK*. MEANS program, Bruselas: DGXVI CEE.

Rossi, P.H., & Freeman, H.E. (1982, 1985, 1989). *Evaluation: A Systematic Approach*. Beverly Hills, CA: Sage.

Rossi, P.H. & Lyall, K.C. (1978). An Overview Evaluation of the NIT Experiment. En T.D. Cook, M.L. DelRosario, K.M. Hennigan, M.M. Mark, & W.M.K. Trochim (Eds.). *Evaluation Studies Review Annual* (Vol. 3, pp. 412-428). Beverly Hills, CA: Sage.

Scriven, M. (1967). The Methodology of Evaluation. En R.W.Tyler, R.M.Gagne, & M.Scriven (Eds.). *Perspectives of Curriculum Evaluation* (pp. 39-83). Chicago: Rand McNally.

— (1972). The Methodology of Evaluation. En C.H. Weiss (Ed.). *Evaluating Action Programs: Readings in Social Action and Education* (pp. 123-136). Boston: Allyn & Bacon.

— (1974). Evaluation Perspectives and Procedures. En J.W.Popham (Ed.), *Evaluation in Education: Current Application* (pp. 3-93). Berkeley, CA: McCutchan.

— (1980). *The Logic of Evaluation*. Inverness, CA: Edgepress.

Scriven, M. (1981). The Good News and the Bad News about Product Evaluation. *Evaluation News*, 2, 278-282.

— (1983). Evaluation Ideologies. En G.F. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation* (pp. 229-260). Boston: Kluwer-Nijhoff.

Shadish, W.R. Epstein, R. (1987). Patterns of Program Evaluation Practice. Among Members of the Evaluation Research Society and Evaluation Network. *Evaluation Review*, 11, 555.

Shadish, W.R. Cook, T.D. Leviton. L.C. (1991). *Foundations of Program Evaluation: Theories of Practice*. Newbury Park, Ca: Sage.

Smith, N.L. (1981). Classic 1960s Articles in Educational Evaluation. *Evaluation and Program Planning*, 4, 177-183.

Stake, R.E. (1975a). An Interview with Robert Stake on Responsive Evaluation. En R.E. Stake (Ed.). *Evaluating the Arts in Education: A Responsive Approach* (pp. 33-38). Columbus, OH: Merrill.

— (1975b). To Evaluate an Arts Program. En R.E. Stake (Ed.). *Evaluating the Arts in Education: A Responsive Approach* (pp. 13-31). Columbus, OH: Merrill.

— (1978). The Case Study Method in Social Inquiry. *Educational Researcher*, 7, 5-8.

— (1979). Validating Representations: The Evaluator's Responsibility. En R. Perloff (Ed.). *Evaluator Intervention: Pros and Cons* (pp. 55-70). Beverly Hills, CA: Sage.

— (1980a). Program Evaluation, Particularly Responsive Evaluation. En W.B. Dockrell & D. Hamilton (Eds.). *Rethinking Educational Research* (pp. 72-87). London: Hodder & Stoughton.

- (1980b). Quality of Education and the Diminution of Local Control in Schools in the United States. En U.S. House of Representatives, Committee on Education and Labor, *Needs of Elementary and Secondary Education in the 1980's* (pp. 161-168). Washington, DC:U.S. Government Printing Office.
- Stake, R.E., & Gjerde, C. (1974). An Evaluation of TCITY: The Twin City Institute for Talented Youth (*AERA Monograph Series in Curriculum Evaluation*, N.º 7). Chicago: Rand McNally.
- Stake, R.E. & Easley, J.A. (1978). *Case Studies in Science Education*. Champaign. University of Illinois, Center for Instructional Research and Curriculum Evaluation.
- Weiss, C.H. (1972a). Evaluating Educational and Social Action programs. A Treeful of Owls. En C.H. Weiss (Ed.). *Evaluating Action Programs. Readings in Social Action and Education* (pp. 3-27). Boston: Allyn & Bacon.
- (1972b). *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, N. J: Prentice-Hall.
- (1973a). The Politics of Impact Measurement. *Policy Studies Journal*, 1, 179-183.
- Weiss, C.H. (1973b). Where Politics and Evaluation Research meet. *Evaluation*, 1, 37-45.
- Weiss, C.H. (1980). Efforts at Bureaucratic Reform: What Have We Learned? En C.H. Weiss & A.H. Barton (Eds.) *Making Bureaucracies Work* (pp. 7-26). Beverly Hills, CA: Sage.
- Weiss, C.H. (1983a). Toward the Future of Stakeholder Approaches in Evaluation. En A.S. Bryk (ed.), *Stakeholder-Based Evaluation* (pp. 83-96). San Francisco: Jossey-Bass.
- Weiss, C.H. (1983b). Ideology, Interest, and Information: The Basis of Policy Decisions. En D. Callahan & B.Jennings (Eds.), *Ethics, the Social Sciences, and Policy Analysis* (pp. 213- 245). Nueva York: Plenum.
- Weiss, C.H. (1987). Evaluating Social Programs: What Have We Learned? *Society*, 25 (1), 40-45.
- Weiss, C.H. & Bucuvalas, M.J. (1980). *Social Science Research and Decision-Making*. Nueva York: Columbia University Press.
- Weiss, R.S., & Rein, M. (1970). The Evaluation of Broad-Aim Programs: Experimental Design, its Difficulties, and an Alternative. *Administrative Science Quarterly*, 15, 97-109.
- Wholey, J.S. (1979). *Evaluation: Promise and Performance*. Washington, DC: Urban Institute.
- Wholey, J.S. (1983). *Evaluation and Effective Public Management*. Boston: Little Brown.
- Wholey, J.S., Abramson, M.A. y Bellavita, C. (eds.) (1986). *Performance and Credibility. Developing Excellence in Public and Nonprofit Organizations*. Lexington, MA: Lexington.

Investigación de la evaluación*

Evert Vedung

En su sentido cotidiano, la evaluación alude al proceso general de determinar el mérito, la importancia y el valor de algo —o el producto de ese proceso—. Implica distinguir lo valedero de lo inútil, lo precioso de lo falto de valor. Sin embargo, en la actual administración del sector público, la evaluación ha adquirido significados más específicos y limitados. Allí la evaluación es un mecanismo para monitorear, sistematizar y calificar las intervenciones gubernamentales en curso o recientemente finalizadas (políticas, programas, proyectos, actividades y sus efectos, y los procesos anteriores a estos efectos, incluidas las percepciones del contenido de la intervención), de modo que los funcionarios públicos y las partes interesadas puedan actuar de la manera más responsable, creativa, equitativa y económica posible en su trabajo orientado al futuro.

En el presente artículo, se adoptará la siguiente definición de evaluación: la cuidadosa valoración del mérito, importancia o valor del contenido, administración, producción y efectos de las intervenciones gubernamentales en curso o finalizadas, que se pretende desempeñe un papel en futuras situaciones y acciones prácticas.

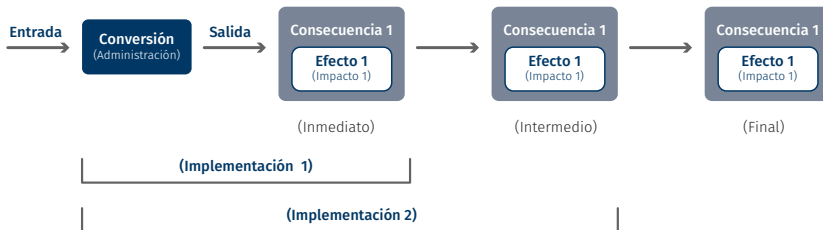
La evaluación se trata de recolección y análisis de datos, pero también de aplicación de criterios de valor a esos datos recolectados y analizados.

La visión de sistema

Los evaluadores tienden a considerar al sector público como un sistema. Un sistema es un todo cuyas partes dependen las unas de las otras. En su versión más simple, un sistema está formado por una entrada, un proceso de conversión y una salida.

Cuando se utiliza el modelo de sistema simple en la evaluación del sector público, se agregan más funciones y la terminología cambia. La etapa de *conversión* pasa a llamarse «administración» y la fase de efectos vira a la etapa de producción. Cuando nos referimos a *producción*, aludimos a los fenómenos que surgen de los órganos gubernamentales, como prohibiciones, concesiones, subsidios, impuestos, exhortaciones, invocaciones morales, bienes y servicios. La consecuencia es lo que sucede con los destinatarios, incluidas sus acciones, pero también lo que ocurre en la cadena de influencia. Podemos distinguir entre consecuencias inmediatas, intermedias y finales. Los *efectos* son un subgrupo de consecuencias, es decir, la porción de resultados producidos, al menos en cierta medida mínima, por la intervención y por las actividades de intervención. Otro término que también significa efectos es «impactos». La palabra «resultados» se usa como un término que condensa tanto a productos, efectos o consecuencias, como a los tres juntos. En general, el término «implementación» abarca la administración y la producción. Este razonamiento se resume en la Figura 1.

Figura 1. El modelo general del sistema adaptado a la evaluación de la intervención gubernamental



Fuente: Adaptado de Vedung, 1997: 5.

El enfoque de los ocho problemas para la evaluación

Los problemas que se atacan en la evaluación pueden formularse como ocho preguntas. Lo llamaré «el enfoque de los ocho problemas para la evaluación de políticas públicas».

1. El problema del propósito integral: ¿para qué objetivos globales se pone en marcha la evaluación?
2. El problema de la organización (evaluador): ¿quién debería realizar la evaluación y cómo debería organizarse?
3. El problema de los criterios de valor: ¿con qué criterios se pueden o se deberían valorar los méritos de la intervención? ¿Según qué estándares de desempeño basados en criterios de valor puede juzgarse el éxito, el fracaso o el rendimiento satisfactorio? ¿Cuáles son los méritos y deméritos reales de la intervención?
4. El problema del análisis de la intervención: ¿cómo se debe caracterizar y describir el objeto de la evaluación, es decir, la intervención gubernamental, normalmente, la política, programa, componentes de políticas y programas, proyecto, la oferta de bienes y servicios o la estrategia de gestión?
5. El problema de implementación: ¿hay obstáculos o fallas en la fase de ejecución desde la instauración formal de la intervención hasta los resultados finales? ¿Cómo pueden mitigarse estos problemas?
6. El problema de las consecuencias: ¿cuáles son las posibles consecuencias relevantes —inmediatas, intermedias y finales, intencionadas y no intencionadas— de la intervención?
7. El problema del impacto (efecto): ¿hasta qué punto son importantes los efectos del resultado de la intervención? Además de la intervención, ¿qué otras contingencias (fuerzas causales operativas) o factores (mecanismos) contribuyeron al resultado?
8. El problema de utilización: ¿cómo debe utilizarse la evaluación? ¿Cómo se usa realmente? ¿Cómo se puede mejorar la utilización?

Propósitos integrales de la evaluación

El fundamento general de la evaluación es crear repositorios de discernimientos descriptivos y críticos para el pensamiento práctico sistematizado y para la acción. Bajo el paraguas de la prestación de servicios, la evaluación se realiza tanto para la rendición de cuentas, la mejora o el discernimiento básico. Además, la evaluación está atada a objetivos estratégicos para ganar tiempo o para mostrarle al mundo un frente de racionalidad.

El fundamento clave de la evaluación para la *rendición de cuentas* es averiguar cómo los agentes ejercieron sus facultades delegadas, de manera que los principales puedan juzgar su trabajo y adoptar la acción correspondiente. La primera dimensión de la rendición de cuentas es presentar informes en los que se proporcionan descripciones. Una segunda función es un análisis justificativo o explicación. En los casos en los que el análisis justificativo reconoce deficiencias, la verdadera rendición exige responsabilización —es decir, quienes están a cargo deben ser garantes de sus actos—.

En la perspectiva de *mejoramiento*, la evaluación aspira a orientar la mejora de la intervención en su implementación. Se estima que la intervención continuará sus operaciones en un futuro previsible y, dentro de lo posible, debe funcionar sin problemas. En esta categoría, los subobjetivos incluyen la eficacia, la rentabilidad, la sostenibilidad, la equidad y la adaptación a las preocupaciones y necesidades del cliente.

El principal usuario esperado de la evaluación de mejora es el personal de la organización y la gestión inmediatamente responsable de la intervención y que está involucrado en ella. La evaluación para la rendición de cuentas es una herramienta para que las máximas autoridades controlen a sus subordinados y los hagan responsables tanto de sus acciones como de la intervención.

La mejora de la intervención, o «aprendizaje», «aprendizaje de la experiencia», «evaluación formativa», «promoción», es una aspiración encomiable de la evaluación. Varios expertos, en especial, en el ámbito de

la educación, sostienen que el aprendizaje debe ser el principal propósito integral de la evaluación.

En el tercer propósito integral, *discernimiento básico*, la evaluación es considerada un tipo de investigación fundamental que busca aumentar la comprensión general de las acciones y de los eventos en el sector público. Probablemente, este sea el principal objetivo de las investigaciones evaluativas que los académicos inician, llevan adelante y financian. En las evaluaciones encargadas a los agentes del sector público, como deben ser rápidamente informadas, este propósito es secundario a la rendición de cuentas y al mejoramiento, y son mejor vistas como una consideración derivada del cumplimiento de las otras dos. En la metaevaluación o en la síntesis de la evaluación, la ambición básica de conocimientos profundos se mantiene firme.

Puesto que la evaluación de la intervención pública acontece en escenarios de acción, siempre está impregnada de propósitos *estratégicos*, a menudo ocultos. Los agentes urden utilizar la evaluación con el fin de ganar tiempo, ocultar defectos a sus principales, mostrar atractivas imágenes de programas y proyectos y, en general, dar una apariencia más favorecedora que la realidad. Las evaluaciones se lanzan para convertirse en pueblos de Potiomkin.

El problema de la organización (evaluador)

¿Quién debería realizar la evaluación y cómo debería organizarse? Una opción es la evaluación interna versus la externa.

La *evaluación interna* es el uso que una organización hace de evaluadores que son empleados de esta con el fin de valorar sus propias intervenciones. Está coordinada y financiada por la organización, es llevada adelante por personas que pertenecen a ella, se centra en sus propias actividades y es principalmente para uso de sus integrantes.

Cuando las actividades de una organización son evaluadas por los empleados de una entidad externa, entonces la *evaluación* es *externa*. Esta puede ser coordinada y utilizada por empleados de la organización, pero,

por definición, debe ser producida por personas ajenas a esta. Cuando un principal encarga a una agencia subordinada la confección de un informe de rendición de cuentas y esta contrata a una consultora para realizar el estudio, la valoración es externa.

Las evaluaciones externas pueden ser subcontratadas a consultoras con fines de lucro o a instituciones sin fines de lucro. Estas últimas incluyen las comisiones públicas de investigación, agencias gubernamentales de auditoría, universidades y otras instituciones de estudios superiores, institutos de políticas públicas financiados por el gobierno y usinas de ideas y fundaciones sin fines de lucro. Los medios de comunicación, las usinas de ideas y las universidades también pueden realizar evaluaciones por iniciativa propia y con su financiación.

Criterios de mérito

Un proceso clave de la evaluación es determinar el mérito, el valor o la calidad de la intervención pública sometida a evaluación. El dilema está en qué constituye una intervención pública valiosa y cómo puede estimarse. Esto queda plasmado en estos cuatro pasos:

- Identificación de los criterios de mérito adecuados para la valoración.
- Con respecto a los criterios elegidos, selección de estándares de desempeño que constituyen éxito o fracaso.
- Establecimiento del desempeño real del objeto de la evaluación para cada criterio y comparación con cada estándar.
- Decisión de integrar o no los juicios en una única valoración general de la valía de la intervención.

Los criterios se pueden determinar al comienzo de la evaluación (*ex ante*), durante el proceso (*ex nunc*) o una vez terminada (*ex post*).

Además, los criterios pueden ser descriptivos o prescriptivos. En la *valoración descriptiva*, el evaluador elige los valores de otros como criterios y estándares. En la *valoración prescriptiva*, el mismo evaluador aboga por la primacía de valores particulares, como justicia, equidad y necesidades

del cliente, sin importar si estos son adoptados por el órgano decisor o sostenidos por algún representante de las partes interesadas.

Aparte de las orientaciones generales que deben adoptarse (*ex ante-ex nunc-ex post*, descriptiva-prescriptiva), deben justificarse los valores particulares preferidos en cada orientación. Los criterios sustantivos, económicos y de proceso más comúnmente utilizados se detallan en la Figura 2.

Figura 2. Criterios de mérito sustantivos, económicos y de proceso en evaluación

I Criterios sustantivos de mérito

1. Metas (evaluación eficaz).
2. Metas más criterios para valorar efectos secundarios.
3. Criterios del cliente.
4. Preocupaciones y problemas de las partes interesadas.
5. Criterios profesionales: criterios de pares, criterios propios.
6. Problema subyacente.

II Criterios económicos de mérito

1. Economía (solo el costo).
2. Productividad.
3. Eficiencia: rentabilidad, costo-beneficio.

III Criterios de proceso de mérito

1. Legalidad.
 2. Equidad (imperio de la ley).
 3. Publicidad.
 4. Representatividad.
 5. Democracia participativa (participación pública en la toma de decisiones final).
 6. Democracia deliberativa.
-

Los criterios sustantivos abordan principalmente el contenido, los productos y los efectos sustantivos de la intervención (y secundariamente, los procesos anteriores a ellos y que los conectan). Los criterios económicos se ocupan solo de los costos de la intervención o de la relación entre costos de la intervención y resultados sustantivos y de impacto. Por último, los criterios de proceso verifican la legalidad, equidad, publicidad, representatividad, democracia participativa, democracia deliberativa y otras cualidades de los procedimientos de acuerdo con los cuales los ministerios y las agencias deberían manejar las intervenciones.

Metas

El uso de metas y objetivos institucionalizados de la intervención es un clásico en evaluación. La literatura previa daba por sentado que los evaluadores debían usar metas prefijadas de intervención como criterios de valor. En la actualidad, se aplican muchos más criterios.

No obstante, hay diversas y valederas *razones a favor de la valoración para el logro de metas* en el sector público. El argumento de la democracia representativa es primordial. En una democracia, todo el poder pertenece al pueblo. Aun así, este no puede tomar todas las decisiones sobre los ciudadanos y la naturaleza por falta de competencia y de tiempo. El pueblo no tiene ni el tiempo ni la capacidad necesarios para participar en millones de decisiones sobre, por ejemplo, la asignación de pacientes en listas de espera quirúrgicas o el cuidado diario de adultos mayores enfermos en hogares del sector público. Por estos motivos, la ciudadanía debe elegir representantes políticos para que tomen decisiones por ellos. Sin embargo, los representantes en las asambleas políticas no tienen ni el tiempo ni la capacidad para tomar todas las decisiones. Por consiguiente, deben delegar sus facultades a los gobiernos para que decidan por ellos. Pero los gobiernos no tienen ni tiempo ni el conocimiento crucial, así que, a su vez, deben delegar sus facultades a los funcionarios civiles y profesionales para que tomen las decisiones, y así sucesivamente. El sector público está formado por largas cadenas de relaciones de principal-agente.

Para salvaguardar que los agentes no engañen a sus principales, el establecimiento de metas del sector público está limitado por sistemas normativos formales. Los responsables de determinar las metas están obligados a respetar la Constitución y las normas de procedimiento específicas. Una vez que una decisión en materia de metas sale del sistema, adquiere un estatus institucional especial.

Esto es lo que sucede con todas las entidades del sector público. Si una agencia adopta un programa con el fin de alcanzar ciertas metas, estos objetivos establecidos están legitimados por el hecho de que el gobierno delegó la autoridad para la toma de decisiones en la agencia y de que, a su vez, el gobierno recibió la facultad para hacerlo del Parlamento y este, a su

vez, del pueblo. Es un mérito del enfoque de logro de metas que reconoce este aspecto democrático de las metas del sector público.

Asimismo, este enfoque también padece debilidades. El *argumento de vaguedad* sostiene que las metas predeterminadas de la intervención son deficientes como criterios de mérito debido a su oscuridad. En ocasiones, los programas se basan en *metas indeterminadas*. Las metas particulares pueden ser *ambiguas* y pueden tener dos o más significados simultáneos. Aun así, la ambigüedad en este sentido de doble significado es excepcional en el lenguaje político y burocrático, por lo que apenas preocupa a los evaluadores. La *vaguedad* causa más incertidumbre. El margen externo que delimita el alcance de una palabra vaga es tan confuso que, en un cierto rango, resulta imposible saber qué se incluye en ese alcance y qué no.

Incluso así, el problema real con la vaguedad es creado por el *catálogo de metas*. La mayoría de las grandes reformas sociales contienen impresionantes directorios de metas diferentes. Si bien una única meta puede ser proclamada como la principal, en general, esta debe sopesarse con todo el resto, tal vez, incluyendo metas potencialmente rivales. Sin embargo, no se indican las compensaciones entre las diversas metas, lo que imposibilita obtener una consecuencia distinta, transparente y esperada de esas listas de metas. De este modo, una serie de metas del programa de este tipo no es lo bastante lúcida para ser empleada como criterios de valor contra la que medir éxitos y fracasos de la intervención.

La esencia del *argumento inflacionario* es que se exageran las metas. Se establecen no solo para alcanzarlas, sino también para conseguir el respaldo de diversas audiencias. Por lo tanto, las metas no son buenos criterios de valor para el desempeño y el logro.

La refutación más convincente del enfoque de metas predeterminadas surge de su ceguera frente a los *efectos secundarios no intencionados*. Las intervenciones del sector público invariablemente provocan consecuencias que no se previeron en la situación original de la decisión. Dado que nadie puede establecer por adelantado metas razonables para algo desconocido, el evaluador guiado por metas premeditadas establecidas sencillamente no podrá rastrear ningún efecto imprevisto. Y si llega a descubrirlo, no

habrá metas preestablecidas que puedan funcionar como criterios de valor para juzgarlas. Sin duda, el enfoque de metas preestablecidas es deficiente en esta cuestión.

Criterios para los efectos secundarios

Es tarea obligatoria de la evaluación mapear y valorar los resultados de la serendipidad y los efectos secundarios no anticipados fuera del área objetivo de las metas. El término «efectos secundarios» se refiere a las implicancias al menos parciales de la intervención que suceden fuera de sus áreas objetivo. Al tomar esto en consideración, los evaluadores deberían siempre buscar los efectos secundarios, al mismo tiempo que, por supuesto, no ignoran los efectos principales. En estas iniciativas, las metas preestablecidas de intervención son conservadas como patrones de valor fundamentales para los efectos principales. No obstante, para los efectos secundarios no anticipados, los criterios de valor deben desarrollarse o en el proceso de evaluación o cuando finalizó la evaluación (*ex post*).

Criterios orientados al cliente

La receptividad a las *preocupaciones de los clientes de la intervención* ha sido planteada como una alternativa a las metas de la intervención y a los criterios de mérito para los efectos secundarios, o como un complemento de ellos. Algunas causas de estas preocupaciones son los deseos, los valores, los objetivos y las expectativas de los destinatarios de la intervención. En el centro de un enfoque orientado al cliente, se encuentra la cuestión de si la intervención, su implementación, los productos y los efectos satisfacen los deseos, valores e intereses del cliente, o si son acordes con las expectativas, suposiciones y necesidades del cliente. Los criterios del cliente pueden incluirse de manera débil, moderada o fuerte en la evaluación. En el caso de incluirlos débilmente, los participantes completan los instrumentos de recolección de datos y brindan información, pero nada más. Los estudios de satisfacción del cliente sobre servicios del sector público pertenecen a esta categoría.

El hecho de incluirlos fuertemente supone que la evaluación es iniciada, financiada, diseñada, realizada e informada por los propios clientes.

Permítanme razonar a partir de un caso de inclusión moderada en el que la evaluación (1) es encargada por los administradores, pero (2) establecida para que involucre a los usuarios de los servicios en las decisiones importantes sobre su planificación y ejecución. Una vez que se designa un equipo de evaluación y se acuerdan sus términos de referencia, se localizan los clientes de la intervención y se realiza un muestreo. Acto seguido, se solicita a los clientes que elijan las dimensiones de intervención que se juzgarán y los criterios de valor que se usarán. A modo de ejemplo, se les pide que juzguen la producción de un programa, la disponibilidad de un servicio, su calidad e incluso su proceso y administración. ¿Se adapta el servicio principal a las demandas del cliente? ¿Los empleados se acercan al cliente con respeto y corrección? Estas son dos preguntas que los usuarios del servicio podrían responder en sus autoevaluaciones.

Los clientes también pueden elegir plantear la cuestión de los efectos, en otras palabras, estimar los impactos del servicio en sus propias vidas o en la comunidad de clientes en general. En dicha evaluación, los destinatarios tratan de determinar el cambio relativo en ellos mismos o en el conjunto de clientes como resultado de su participación en modalidades específicas de tratamiento.

En la recolección de datos, los defensores de los criterios orientados al cliente prefieren la autoobservación de los clientes y la realización continua de entrevistas a los cuestionarios, los métodos documentales y la observación del evaluador. Las entrevistas en profundidad de los destinatarios individuales es una de las técnicas preferidas, es decir, la distribución de instrumentos de autoevaluación que los clientes y su red pueden fácilmente completar. En algunos casos, los evaluadores orientados al cliente defienden las *entrevistas de grupos focales* que permiten las *deliberaciones grupales* entre participantes, y entre estos y el evaluador. Estos foros de debate respaldarán el desarrollo de nuevas ideas, conceptos de servicios, soluciones y tecnologías. Como un efecto secundario, también podrían educar a los participantes para que se conviertan en mejores ciudadanos en el futuro.

Hoy los criterios de valores basados en el cliente se usan en numerosos contextos evaluativos, particularmente en aquellos sobre la provisión

de servicios públicos, como la atención de la infancia, los hogares para adultos mayores, la vivienda pública, la salud mental, el tránsito urbano, los servicios públicos, las plazas y la recreación, así como los servicios de salud física, donde la participación de la clientela es crucial para el funcionamiento del programa. Los criterios basados en el cliente se utilizan para evaluar los servicios de bibliotecas, los museos, los parques nacionales, las piscinas, las canchas de fútbol, las canchas de tenis, la recolección de basura, la limpieza de las calles, la remoción de nieve, el ruido del tránsito, la congestión vehicular y el tránsito urbano. Es la opción preferida de los educadores. En las universidades, siempre se les pide a los estudiantes que compartan sus opiniones sobre los cursos, listas de lectura y conferencias. También, que califiquen las aptitudes de sus profesores para organizar los contenidos del curso, estimular y promover la discusión y el debate, suscitar el interés de los alumnos, promover el pensamiento crítico y mostrar preocupación y entusiasmo por los alumnos. En algunos contextos, estas evaluaciones se usan ocasionalmente para calificar al cuerpo docente y los cursos, de manera que los futuros estudiantes puedan tomar decisiones más informadas.

El uso de las preocupaciones de los clientes como criterios de valor ajustará las intervenciones públicas para servir a los clientes. Se basa en la noción de que la administración pública produce bienes y servicios para los clientes en el mercado. Al comprar un *commodity* en una tienda, el cliente no presta atención a las metas del productor. Su propia valoración del valor del bien es lo único que importa. Los servicios públicos también deberían ser considerados de esta forma y estar orientados a los gustos de los consumidores. De hecho, el principal fundamento de los servicios públicos es servir a los clientes. Por consiguiente, solo importa la propia valoración del cliente del valor del bien. En situaciones en las que los destinatarios están involucrados moderada o fuertemente en la evaluación, se crea un argumento adicional en materia de *democracia participativa* y de *democracia deliberativa* para el uso de los criterios basados en el cliente. De acuerdo con esta justificación, la analogía del cliente no puede forzarse demasiado, ya que la noción del cliente incluye tanto un aspecto participativo como uno deliberativo que está ausente en el concepto de consumidor. La característica participativa sugiere que los clientes también son ciudadanos que pueden expresar quejas y deseos a los evaluadores y a los proveedores

de servicios, e influenciar, en cierta medida, el contenido del servicio y asumir la responsabilidad por este. La característica deliberativa genera un consentimiento discursivo, racional, de debate y de aprendizaje a través del diálogo que puede educar a los clientes para que sean mejores ciudadanos en general: el cliente como ciudadano y no como consumidor. Aun así, la principal idea es que los servicios estén más orientados a los deseos del consumidor. El aspecto vinculado con el hecho de ser mejores ciudadanos es considerado un efecto secundario.

La adopción de criterios orientados al cliente también puede aumentar la *legitimidad* de la intervención. Si los clientes pueden participar y ejercer cierta influencia en la intervención, probablemente, aumentará su aprobación de esta.

Asimismo, la aplicación de criterios orientados al cliente puede fomentar la *eficacia* y la *eficiencia*, ya que concentrarse en los clientes forzará a proveedores y administradores a eliminar otras cuestiones que les preocupan además de brindar buenos servicios.

No obstante, los evaluadores deben estar conscientes de la tendencia de la clientela a exagerar las quejas para obtener más servicio. Los clientes también pueden alimentar ilusiones fiscales. Una mayor participación de los clientes en la evaluación puede significar la entrega del poder a grupos con intereses creados y limitados.

En definitiva, la aplicación de criterios orientados al cliente puede complementar los criterios de mérito antes presentados, dado que ponen en consideración otros problemas. Pueden hacer importantes contribuciones a la evaluación, pero deben sopesarse con otros criterios, como el logro de metas y la normativa profesionales para la excelencia del servicio.

Modelos profesionales: autoevaluación más revisión por pares

Otro conjunto de criterios de valor son las *concepciones profesionales del mérito*. A los miembros de una profesión se les confía la evaluación de su propio desempeño y el de sus colegas en relación con los criterios de mérito y los estándares de calidad de la profesión.

En algunas áreas de la vida pública, los criterios de calidad son tan complejos y el objeto de estudio es tan intrincado que los funcionarios consideraron acertado dejar que profesionales bien formados debatieran y moldearan estos criterios. Los arquitectos, los jueces, los profesores universitarios, los médicos, las enfermeras, los veterinarios y los ingenieros serían algunos ejemplos de ello. En consecuencia, también se considera natural delegarles la evaluación *ex nunc* y *ex post* a las profesiones.

La evaluación de la investigación de un departamento de economía de una universidad europea puede usarse como ejemplo. A menudo, la evaluación profesional comienza con una autoevaluación de los evaluados. Los profesionales que serán evaluados realizan una valoración de su propio desempeño. ¿Cuáles son las fortalezas y las debilidades de nuestro departamento? ¿Qué oportunidades tenemos? ¿A qué amenazas debemos enfrentarnos ahora o en un futuro cercano? En resumen, realizan un *análisis FODA*.

Después viene una *revisión por pares*. Científicos muy renombrados de un campo de estudio en particular, pero de otras universidades, son designados para valorar la calidad y la relevancia de los académicos. Además de la autoevaluación escrita, estos pares basan su valoración en evidencia documental adicional y en visitas a los diferentes lugares para mantener diálogos cara a cara entre evaluador y evaluado. Los evaluadores vuelcan sus juicios en un informe preliminar, el cual suele estar organizado de acuerdo con el esquema FODA. Posteriormente, los evaluados tienen la oportunidad de hacer comentarios sobre este informe preliminar. El objetivo particular del informe final de la revisión por pares es realizar un juicio general de calidad del objeto de la evaluación.

A menudo, la evaluación basada en criterios profesionales de mérito es interactiva. Los evaluadores se reúnen con los evaluados y los escuchan, y los evaluados aprenden de los evaluadores. Son importantes el diálogo, la deliberación y la interacción entre los evaluados, y entre estos y los evaluadores.

Las autoevaluaciones y la revisión por pares frecuentemente producen resultados cuestionables. Los estudios con análisis de paneles coincidentes

muestran que los pares usan criterios de mérito y estándares de desempeño muy disímiles, por lo que llegan a conclusiones diferentes. Sin embargo, en los campos complejos, es probable que la evaluación colegiada interactiva sea el mejor método disponible para juzgar la calidad de lo que se produce.

Preocupaciones y problemas de las partes interesadas

Por último, las preocupaciones y los problemas de todos los actores que tienen un interés en la intervención o que se ven afectados por ella pueden usarse como patrones de evaluación.

Las partes interesadas pueden constituirse como equipo de evaluación y llevarla adelante. A continuación, quiero razonar a partir de otro caso en el que los evaluadores particulares conducen la evaluación y generan los puntos de vista de las partes interesadas.

El enfoque de las partes interesadas comienza cuando el evaluador mapea los principales grupos involucrados o que podrían tener interés en la intervención. El evaluador identifica a las personas que iniciaron, trabajaron duro, financiaron y adoptaron la intervención. Además, identifica a quienes están a cargo de su implementación: los administradores, el personal y los operadores de primera línea.

El evaluador señala al principal grupo destinatario de la intervención, los clientes, y a las asociaciones de los clientes. Identifica a los familiares y a sus asociaciones. También puede incluir legos. Asimismo, busca a quienes no están conscientes del interés que sostienen.

Los defensores de los criterios basados en las partes interesadas fomentan una fuerte predilección por la metodología cualitativa, interactiva y práctica. El evaluador debe hablar con las partes interesadas para obtener sus narrativas y datos observacionales. El evaluador debe ser receptivo y dejar que los problemas y las preocupaciones de las personas afectadas guíen el próximo paso de la iniciativa de investigación. Después de un tiempo, puede descubrir qué problemas ocupan las mentes de las diversas partes interesadas y qué preocupaciones alimentan respecto de estos problemas. El diseño de la

evaluación se determinará gradualmente. La evaluación de las partes interesadas es una *evaluación receptiva*.

Para recopilar los datos finales sobre los problemas y las preocupaciones de las partes interesadas, se prefieren la autoobservación de las partes interesadas y los métodos dialogales constantes a las encuestas, los cuestionarios y las estadísticas. Una de las técnicas favoritas es la entrevista en profundidad de los destinatarios individuales. Se usan instrumentos de autorreporte que los clientes, los padres, los familiares y otras redes de partes interesadas pueden completar con facilidad. En algunos casos, los evaluadores de las partes interesadas defienden las *entrevistas de grupos focales*, que permiten las *deliberaciones grupales* entre participantes, y entre estos y el evaluador. Después de la acumulación y el procesamiento de los datos, comenzará la presentación de informes sobre los hallazgos, que pueden variar de una parte interesada a la otra. La palabra clave parece ser «representación», es decir, caracterizaciones ricas en información usando imágenes, anécdotas, abundantes descripciones y citas.

El uso de criterios basados en las partes interesadas tiene varias ventajas. Los *argumentos democráticos* se apartan de los puntos de vista participativos y deliberativos. Es cierto que la democracia supone que en las elecciones generales, los ciudadanos votan a élites rivales que toman las decisiones en su nombre (*democracia representativa*). Incluso así, la ciudadanía también debería poder participar en la toma de decisiones pública final entre elecciones (*democracia participativa*). Además, la discusión, el diálogo y el debate también son valores democráticos importantes porque ayudan a las personas a formar y perfeccionar sus creencias y preferencias (*democracia deliberativa*).

Los modelos basados en las partes interesadas satisfacen más los valores participativos y deliberativos que los enfoques de logro de metas y de efectos secundarios. También, y como externalidad positiva, puede educarse a los participantes para que se conviertan en mejores ciudadanos en el futuro.

De acuerdo con el *argumento del conocimiento*, las partes interesadas alimentan convicciones sobre los efectos secundarios no intencionados,

las sofisticadas barreras de implementación y el engaño absoluto que pueden proporcionar al investigador ideas sobre tópicos para mayores investigaciones. Puesto que la orientación concerniente a las partes interesadas traerá más aspectos del objeto de estudio en discusión y aumentará la calidad de los hallazgos de la evaluación.

También existe un *argumento de utilización* que respalda la evaluación de las partes interesadas. El enfoque orientado a las partes interesadas incrementa las chances de que se aborden temas de genuino interés para las partes en cuestión. Saca a la luz información que cumple con los requerimientos reales de las distintas partes interesadas y, por consiguiente, mejora la probabilidad de que se haga uso de los hallazgos.

Un último fundamento a favor de los modelos basados en las partes interesadas es que pueden promover *compromisos* y prevenir amargas luchas políticas. Las asambleas de las partes interesadas son vehículos para moldear acuerdos sobre los resultados de los primeros esfuerzos y, más importante aún, propuestas para las futuras acciones. Se considera que la creación de consenso y el otorgamiento de legitimidad para las decisiones fundamentales son enormes ventajas de la evaluación de las partes interesadas.

Los criterios basados en las partes interesadas también presentan *desventajas*. Las evaluaciones de las partes interesadas son *desproporcionadamente inviables y demandan recursos*, ya que se debe contactar y estimular a cada representante de las partes interesadas. Los modelos basados en las partes interesadas son *confusos*. No proporcionan ninguna respuesta clara a la pregunta de quiénes son las partes interesadas. Asimismo, todas las audiencias de las partes interesadas, aunque se seleccionan, son tratadas como si fueran iguales. Sin embargo, en una democracia representativa, los políticos elegidos deben cargar con un peso mayor que los administradores o los expertos con respecto a los temas sustantivos en consideración. El modelo basado en las partes interesadas no incorpora *prioridades entre ellas*. Existe el riesgo de que se consulte y se escuche más a las partes interesadas pequeñas, bien organizadas y muy comprometidas que a los grupos vagamente preocupados y a la mayoría de los ciudadanos.

Esto suscita un problema con la evaluación de las partes interesadas, o más precisamente, con la toma de decisiones colectiva sobre la base de discernimientos propuestos por las propias evaluaciones de las partes interesadas. Puesto que la democracia representativa es la forma de gobierno democrático dominante en la mayoría de las sociedades desarrolladas, las decisiones sobre diversas políticas sectoriales no pueden dejarse en manos de las partes interesadas. El poder le pertenece al pueblo, a sus representantes y a sus delegados. La evaluación de las partes interesadas debe trabajar dentro de los marcos fijados por la democracia representativa.

Otra objeción involucra el riesgo de que el enfoque basado en las partes interesadas adopte una *teoría pragmática de la verdad*. La verdad puede convertirse en una cuestión de conveniencia, utilidad o aceptabilidad para las partes interesadas. Por lo general, estas albergan visiones politizadas de los efectos del programa. Los defensores le atribuyen todo lo positivo que sucede posteriormente al programa y todo lo negativo a algo más. Los detractores sostienen lo contrario. Básicamente, los hechos son controvertidos. En estas situaciones, las diversas partes aceptarán solo las conclusiones que se adecuen a sus opiniones preconcebidas.

Costo de los criterios de mérito

Los criterios sustantivos no prestan atención a los costos. Para remediar esta omisión, los economistas elaboraron tres criterios de costos: economía, productividad y eficiencia. La *economía* es un criterio basado únicamente en el costo. ¿La intervención se lleva a cabo de una forma razonablemente barata? ¿Se prefirieron las opciones más baratas respecto de las más caras? ¿La implementación se realiza de la manera más barata posible?

La *productividad* es la ratio entre productos y costos, en otras palabras, productos = insumos. La productividad se operacionaliza de modo diferente, debido al tipo de producción y de costo. Un estudio de la productividad en las bibliotecas municipales puede usar la productividad del costo como una medida: cantidad de libros prestados dividida por costos en euros (o en dólares) de prestar esos libros. Como alternativa a la productividad del costo, se puede usar la productividad del trabajo,

que puede ilustrarse mediante la expresión cantidad de libros prestados dividida por cantidad de horas trabajadas.

En ambos casos, la producción se indica en términos físicos, por ejemplo, cantidad de libros prestados. La diferencia es que en el primer caso los costos se indican en términos monetarios, mientras que en el segundo, en cantidad de horas trabajadas, es decir, en entidades físicas. Quizá, debería destacarse que los costos pueden calcularse de ambas formas en la medición de productividad.

Como patrón de producción, la productividad no es una vara de medición ideal para valorar el mérito de las actividades del sector público. Las instituciones públicas pueden hacer las cosas mal, en otras palabras, los productos pueden no generar los efectos deseados. Por consiguiente, la eficiencia es mejor parámetro como patrón para los efectos.

Idealmente, la eficiencia presupone dos cuestiones: primero, el evaluador debe saber que los efectos son producidos, al menos en parte, por la intervención y no por otra cosa. Segundo, el valor de estos efectos debe calcularse usando algunos criterios de valor como las metas de la intervención. Si la eficiencia se mide en un análisis de costo-beneficio, se puede expresar como la ratio entre valor monetarizado de los efectos de la intervención y costos monetarizados, es decir, valor de los efectos de la intervención (en euros): costos (en euros).

Criterios de proceso

Todos los criterios de evaluación de la intervención pública expuestos hasta aquí brindan perspectivas parciales, porque pasan por alto o tratan como dados otros requerimientos que normalmente demandan las actividades del sector público en las democracias contemporáneas como las metas de proceso. El logro de metas, las pretensiones de los clientes, los estándares profesionales y la eficiencia deben balancearse con la *igualdad jurídica*, la legitimidad, la imparcialidad procedimental y la apertura al escrutinio público. Las metas de los efectos sustantivos y económicos también deben contrastarse con los *valores democráticos*, como la participación del cliente y los procesos deliberativos.

La tarea de combinar, por ejemplo, el logro de metas con la economía, la imparcialidad procedimental, la representatividad y la participación del cliente en un único valor general no se puede, al menos por el momento, cumplir con medios científicos. Esa clasificación global necesita de criterio político.

El problema del análisis de la intervención

¿Qué se puede evaluar? Nada. El presente artículo trata sobre la evaluación de la intervención pública. Esta puede adoptar la forma de una política, una reforma, un programa, el sostenimiento de un programa, un proyecto. También podría ser una actividad, como la provisión del servicio de atención de la infancia, o estrategias administrativas, como la gestión orientada a resultados.

Las intervenciones públicas varían mucho. Algunas son amplias y otras extremadamente limitadas. Algunas pueden durar indefinidamente y otras solo unas pocas horas. Algunas son programas regulares que han continuado durante décadas y otras proyectos con fecha establecida de finalización. Algunas tienen un alcance muy local y otras intrarregional, nacional, interregional (Unión Europea) o incluso mundial. Aun así, sin importar si son amplias o limitadas, duraderas o breves, permanentes o provisionales, deben ser descritas cuando se evalúan —el cuarto punto del «enfoque de los ocho problemas para la evaluación»—. Es esencial que estas representaciones no conviertan las intervenciones en demasiado idiosincráticas ni ligadas a la situación.

Una posibilidad es representar el objeto de la evaluación en términos de sus instrumentos políticos. Los instrumentos de las políticas públicas son un conjunto de técnicas mediante las cuales las autoridades del sector público ejercen su poder para intentar producir un cambio social o generar apoyo. Hay tres y solo tres instrumentos básicos a los que los gobiernos han recurrido, clasificación conocida con la expresión popular «zanahorias, garrotos y sermones». Los gobiernos pueden premiarnos por hacer lo que ellos quieren (zanahorias o incentivos económicos), obligarnos a hacerlo (garrotos o regulaciones) o aconsejarnos que deberíamos hacerlo (sermón o información). Al caracterizar las intervenciones con esta terminología

general o más elaborada, las evaluaciones serán más relevantes y sus hallazgos, más usados y atendidos.

Monitoreo calificado

El monitoreo calificado y la valoración de impacto son dos formas clave de la evaluación. El *monitoreo calificado* alcanza la implementación, los productos y las consecuencias (pero no los efectos) —el quinto y sexto problemas de los ocho problemas para la evaluación—.

El monitoreo calificado, comúnmente denominado «evaluación formativa», chequea las diversas etapas de la cadena de implementación por medio de los productos y los resultados en busca de fallas e impedimentos. Quien monitorea busca, descubre y verifica problemas, desanuda causas, libera mecanismos que los facilitan y sugiere soluciones. ¿La intervención se entrega a los clientes de la manera prevista en la decisión formal de intervención? ¿La manera en que se presenta la intervención alcanza a todos los potenciales participantes? Además, el monitoreo calificado también se centra en los procesos anteriores a la entrega. De hecho, el monitoreo calificado es una evaluación de procesos en el sentido de que se controla la totalidad del proceso de implementación desde la adopción de la intervención formal hasta la posible participación de los destinatarios en la entrega para escudriñar posibles problemas. El punto es analizar si, por ejemplo, una decisión de intervención se realiza de acuerdo con lo planeado en los niveles más bajos de autoridad con el fin de corregir errores y omisiones durante la ejecución y la entrega.

El monitoreo simple, en contraposición al monitoreo calificado, solo genera la recolección de datos de algunas variables sin conjugar esta actividad con las intervenciones, los criterios de valor y la búsqueda de problemas. Es cierto que los datos recopilados a través del monitoreo simple pueden usarse para el monitoreo calificado y para la valoración de impacto. Pero, el monitoreo simple en sí mismo no es una evaluación.

Valoración de impacto

La *valoración de impacto* investiga la cuestión de los efectos que la intervención produce en las consecuencias, ya sean inmediatas, intermedias o finales. La valoración de impacto intenta determinar hasta qué punto las consecuencias son generadas por la intervención o por algo más que opera además de la intervención. La valoración del impacto aborda el séptimo problema del «enfoque de los ocho problemas para la evaluación». ¿Las consecuencias, dentro o fuera de las áreas objetivo, son provocadas y moldeadas, al menos indirectamente o hasta cierto punto, por la intervención o son ocasionadas por algo más?

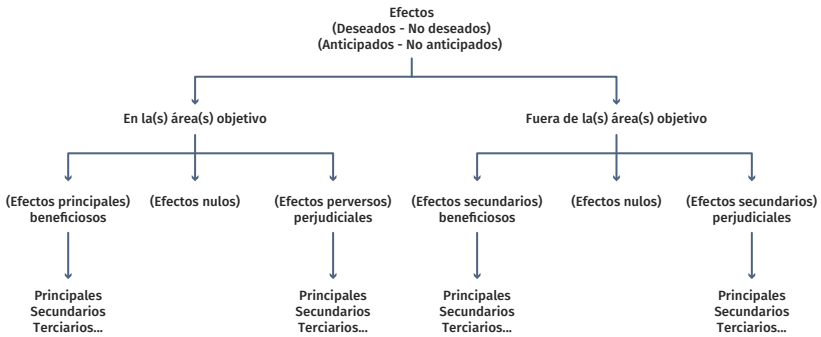
Tipos de efectos

Para capturar los efectos, la comunidad de evaluación ha desarrollado un lenguaje analítico. Se define *efecto principal* como el impacto sustantivo central que los instigadores de la intervención *intencionalmente* quisieron lograr. Los efectos principales se asocian con los objetivos sustantivos de los responsables de la intervención y con lo que esta fue capaz de alcanzar. Además de suceder en el área objetivo, los efectos principales, por definición, son anticipados y valorados positivamente por quienes adoptan la intervención. Del mismo modo, se define *efecto secundario* como una consecuencia que está fuera de las principales áreas objetivo de intervención. Los efectos secundarios pueden anticiparse o no y pueden ser negativos o positivos. Los *efectos perversos* van exactamente en contra de los propósitos establecidos en la intervención. Estos impactos pueden ocurrir en las áreas objetivo de la intervención pública y también pueden surgir fuera de las áreas objetivo y ser efectos secundarios. Los efectos perversos pueden suscitarse muy por debajo de la cadena de control, es decir, en la segunda, tercera o incluso cuarta etapa de las consecuencias. En general, afloran después de muchos años. Los efectos perversos son diferentes de los *efectos nulos*. Estos suponen que las intervenciones no generan ningún impacto dentro o fuera de sus áreas objetivo.

El árbol de efectos de la Figura 3 muestra qué aspectos de los efectos podrían estudiarse en la investigación de la evaluación.

La importancia de darse cuenta de los efectos principales, perversos y nulos debería ser obvia para todos los evaluadores. Pero, ¿por qué resulta vital prestar atención a los efectos secundarios? Porque los derivados, ya sean perjudiciales o beneficiosos, son factores cruciales de cualquier juicio inclusivo del funcionamiento de una intervención.

Figura 3. Efectos principales, secundarios, perversos y nulos



Fuente: Adaptado de Vedung, 1997, 54.

Diseños para la valoración del impacto

No hay una solución extensamente reconocida para el problema del impacto de la evaluación. Se dispone de varios diseños diferentes. A veces, los enfoques pueden tener un orden de clasificación según su capacidad para resolver la cuestión del impacto de la intervención en la evaluación del sector público. La Figura 4 muestra una variante condensada y ampliada de este orden de clasificación, ampliada porque se agregó la evaluación del proceso explicativo a la lista.

Figura 4. Diseños de investigación de la evaluación para el esclarecimiento del impacto

Experimentos controlados aleatorios: en una prueba provisional antes de que se adopte la intervención permanente, se dividen aleatoriamente los destinatarios en un grupo experimental, para el que se administra la intervención, y en un grupo de control —controles aleatorios—, a quienes se les niega la intervención (experimentos clásicos).

Experimentos controlados coincidentes: se comparan los destinatarios de una prueba provisional o quienes estuvieron expuestos a la intervención permanente con un grupo en teoría equivalente, no creado de forma aleatoria a través de la coincidencia —controles coincidentes— a quienes se les niega la intervención o quienes estuvieron expuestos a otras intervenciones (cuasiexperimentos).

Controles genéricos: se comparan los efectos de una intervención provisional o permanente entre los destinatarios con los datos normativos sobre los cambios habituales que suceden en una población que no ha sido objeto de la intervención.

Controles reflexivos: los destinatarios que reciben o que recibieron la intervención provisional o permanente son comparados con ellos mismos, según lo medido antes de la intervención.

Controles estadísticos: se comparan los destinatarios que participan con los que no participan de la intervención provisional o permanente manteniendo estadísticamente constante las diferencias entre los participantes y las personas que la intervención no abarca.

Controles a la sombra: se comparan los destinatarios que reciben o que recibieron la intervención provisional o permanente con los juicios de los expertos, de los administradores del programa y del personal, con los de los propios destinatarios sobre qué cambios piensan que habrían ocurrido si no hubiera habido intervención alguna.

Evaluación del proceso explicativo: se realiza un seguimiento paso por paso del desarrollo de la intervención permanente en su entorno natural, desde la decisión hasta el resultado con el fin de descubrir factores facilitadores y limitantes.

Nota: Una intervención permanente es una intervención «real» en contraste con una intervención provisional.

Fuente: Adaptado de Vedung, 1997: 170.

Los experimentalistas radicales sostienen que los diseños con controles aleatorios o coincidentes constituyen la mejor metodología disponible para solucionar la cuestión de los efectos.

Aun así, para permitir dicha sólida experimentación de dos grupos a mayor escala, la toma de decisiones del sector público debe realizarse en dos etapas bien discernibles. La decisión de instalar intervenciones a escala completa debe estar precedida de pruebas piloto preliminares, diseñadas como experimentos aleatorios o coincidentes. E, igual de importante, los decisores deban considerar el riguroso conocimiento de la evaluación que proviene de estos experimentos cuando, en la segunda etapa, formulan las intervenciones permanentes. De este modo, la política pública se tornará más científica y se basará más en la evidencia y menos en la anécdota y en la improvisación.

Estas son las rigurosas demandas para los procesos de formación de políticas. Por lo general, las condiciones previas —que el grupo destinatario

deba dividirse aleatoriamente en un grupo experimental y otro de control antes del inicio formal de la intervención completa y que el grupo experimental, pero no el de control, debe exponerse a la intervención— no siempre pueden satisfacerse en la política pública.

Asimismo, tanto los experimentos de laboratorio como los de campo *requieren mucho tiempo*, mientras que los organismos políticos a veces deben reaccionar de inmediato. Los experimentos también producen *conocimiento limitado*. En los experimentos casi nunca se bosquejan los procesos entre la adopción de la intervención y sus efectos; los mejores mecanismos de implementación y de respuesta a los destinatarios son tratados como cajas negras. Para los funcionarios públicos, esto representa una desventaja, dado que la información sobre la implementación y sobre las reacciones de los destinatarios es de sustancial interés para ellos.

Aunque los hallazgos experimentales tienen una fuerte *validez interna*, su validez externa es débil. Estos son independientes del contexto, pero en contextos de la vida real, juegan papeles decisivos. Una intervención probada en condiciones experimentales a pequeña escala puede producir efectos totalmente distintos al transferirla a grandes entornos del mundo real.

Por un lado, los evaluadores deberían efectuar sus estudios en estrecho contacto con los evaluados y con las partes interesadas usando, por ejemplo, diálogos, entrevistas y métodos deliberativos. Por otro lado, la experimentación trata a los participantes como conejillos de India e idealmente, estos ni siquiera deberían saber que son parte de un experimento.

El argumento contra los programas públicos experimentales parece ser fuerte. Los investigadores de la evaluación deben sucumbir ante el hecho de que la «parte del león» de las intervenciones gubernamentales ya está en marcha y no puede someterse a ensayos de campo provisionales a pequeña escala. En estos casos, los evaluadores deben trabajar con diseños de investigación de fácil aplicación en la práctica política existente. Se deben enfatizar los controles reflexivos, estadísticos y a la sombra, y, sobre la totalidad de las evaluaciones del proceso de tipo cualitativo, de

estudio de casos y de búsqueda de mecanismos. Además, la evaluación experimental retornó recientemente bajo las formas de medicina y trabajo social basados en la evidencia. El debate continúa.

Análisis de efectos como evaluación de proceso explicativa

En la vida real, la mayoría de los análisis de efectos se efectúan como evaluación de proceso (evaluación del *modus operandi*, seguimiento del proceso), la cual intenta rastrear todos los tipos de consecuencias de la intervención, incluidos los efectos deseados, nulos, perversos y secundarios, ya sean intencionales o no. Adopta un enfoque amplio, configurador y orientado a los mecanismos para los factores explicativos. La evaluación del proceso busca establecer un patrón completo de los mecanismos de interacción entre las intervenciones adoptadas, sus implementaciones, las repuestas de los destinatarios, los contextos y los resultados.

El seguimiento del proceso se concentra en los objetos de evaluación en sus entornos naturales. Se ejecuta en estrecha interacción con los gestores, el personal y los participantes de la intervención. El evaluador busca tener contacto e interactuar con los investigados, no distanciarse ni evitarlos como en la experimentación. La evaluación de proceso no rechaza los datos duros estadísticos, encuestas ni cuestionarios, pero se prefieren las técnicas de recolección de datos cualitativos, como entrevistas en profundidad, grupos focales, revisión documental y observación directa por medio de visitas a los lugares.

Siete factores amplios que podrían influir en los resultados

La cuestión básica de la evaluación explicativa del proceso —vista desde la perspectiva del máximo decisor— puede expresarse de la siguiente forma: ¿en qué modos y hasta qué punto el resultado es producto de la intervención?

En la Figura 5, se organizan siete factores amplios y mecanismos que pueden tener impacto en los resultados de las intervenciones públicas.

Figura 5. Factores explicativos en la evaluación de proceso

I Formación de la intervención

1. Dirección del cambio propuesto.
2. Apoyo político.
3. Política simbólica.
4. Participación de los intereses afectados.

II Diseño de la intervención

1. Claridad: palabras indeterminadas, catálogos de metas.

III Respaldo del decisor después de la instauración**IV Implementación**

1. Agencias nacionales: comprensión, capacidad, buena voluntad (elección pública, actitudes divergentes).
2. Intermediarios formales.
3. Burócratas a nivel de calle o comunitario: estrategias de manejo de situación.

V Respuesta de los destinatarios

1. Comprensión, capacidad, buena voluntad.
2. Momentos formativos.
3. Fanáticos.
4. Captura.
5. *Free-riders*.

VI Otras intervenciones gubernamentales y sistemas de implementación**VII Mercados y otros entornos**

1. Poder del consumidor en el mercado.
2. Apoyo de actores externos a toma de decisiones y administración formal.
3. Medios de comunicación.

Fuente: Adaptado de Vedung, 1997: 212.

Formación de la intervención. La *dirección del cambio propuesto* puede afectar los resultados a través de las consideraciones de los destinatarios o indirectamente mediante las consideraciones realizadas entre los actores de la implementación. Si el cambio previsto apunta hacia una dirección distinta a la que están acostumbrados los funcionarios públicos y los destinatarios, se tornará más difícil llevarlo adelante. Los resultados se verán influenciados por el *apoyo político* para la intervención en relación con su instauración. Los desacuerdos de los grupos partidarios y de presión crean incertidumbre sobre el futuro de un programa, que probablemente desencadene reflexiones y razonamientos, y afecte las decisiones en las agencias nacionales y entre otros actores en la etapa de implementación.

La *política simbólica* puede influenciar en la implementación y en los resultados. Esta supone que la intervención se inicia para otros propósitos

distintos del logro de resultados sustantivos, en otras palabras, para dar la impresión de estar preocupado sin en verdad estarlo. Los responsables de formular las políticas pueden querer satisfacer la opinión del partido o fortalecer su propio liderazgo partidario, mantener a raya a los miembros del partido, asegurar votos en las elecciones generales o facilitar las coaliciones gubernamentales. Naturalmente, los funcionarios de las agencias, los operadores de nivel bajo y otros actores oficiales pueden percibir el contenido simbólico y destinar menos energía a la implementación de la que, de otro modo, hubieran puesto.

La *participación de los intereses afectados* en el trabajo preparatorio que lleva a adoptar la intervención bien puede ejercer una influencia en sus resultados. Si quienes se verán afectados pueden participar en la formación, la intervención obtendrá legitimidad, la que, a su vez, facilitará la implementación.

Diseño de la intervención. La oscuridad de la intervención hace que sea complejo para los funcionarios de la implementación y para los destinatarios crear una imagen correcta de qué políticas están diseñadas para ser alcanzadas. Las *palabras indeterminadas* hacen que el lenguaje de la intervención sea ambiguo e impreciso. Los *catálogos de metas* son una serie de metas sin prioridades entre ellas. En ambos casos, se torna imposible revelar el propósito general de la intervención como fue concebido originalmente. Esto brinda a los funcionarios una amplia discrecionalidad.

Respaldo del decisor después de la instauración. Los responsables de tomar las decisiones pueden respaldar la intervención después de su puesta en marcha, por ejemplo, a través de reiterados apoyos, lo que puede influir en los resultados.

Implementación. La implementación podría afectar los resultados. Se distinguen tres niveles de problemas: agencias nacionales, intermediarios formales y burócratas a nivel de calle o comunitario.

Una *comprensión* deficiente del contenido de la intervención por parte del administrador puede causar problemas de ejecución. El ingrediente

de comprensión es importante en los tres niveles de implementación. Por razones de simplicidad, debo mencionarlo solo en el nivel de la agencia nacional. La *capacidad administrativa* puede modificar los resultados. Si la ejecución de la intervención exige asignaciones, personal, talento, tiempo, energía y equipamientos que no están a disposición de los administradores, la probabilidad de alcanzar resultados exitosos se debilitará. La *buena voluntad* administrativa afecta los resultados. Las burocracias públicas pueden tener sus propias agendas, que pueden entrar en conflicto con la fiel implementación de las directivas y recomendaciones del principal. Una hipótesis muy debatida establece que los burócratas son maximizadores de presupuestos. Con el fin de mejorar su reputación personal, salario, felicidad, confort, esfera de influencia y otros recursos a su servicio, los burócratas tienden a ampliar sus presupuestos.

Una variación más leve del tema de la buena voluntad administrativa es que los funcionarios públicos pueden albergar dudas sobre la pertinencia de la propia intervención, lo que puede entorpecer su disposición para actuar.

Incluso otro subcomponente del factor más grande de la implementación plantea que la comprensión, las capacidades y la buena voluntad de los *intermediarios formales* pueden afectar positiva o negativamente los resultados. Los intermediarios formales ingresan como nexos entre las agencias administrativas nacionales y los grupos destinatarios. Cada vez más se usan intermediarios no gubernamentales para llevar adelante los objetivos de la intervención pública. Una categoría es la de los contratistas comerciales privados y sus organizaciones de interés. Otra es la de las organizaciones de interés público o asociaciones locales voluntarias, como las asociaciones de propietarios, los clubes deportivos o las asociaciones de inquilinos. Los intermediarios no gubernamentales pueden ser socios en la negociación cuando las agencias elaboran normas según un estatuto regulatorio. Pueden estar representados en las juntas de las agencias, pueden actuar como reales implementadores y pueden participar en las evaluaciones. El uso de intermediarios no gubernamentales puede aumentar la relación costo-efectividad y evitar la creación de grandes burocracias. También pueden mejorar la legitimidad del programa a los ojos de los destinatarios y del público general. Aun así, los intermediarios

pueden generar desventajas, ya que la intervención puede desviarse en direcciones que no se ajustan a sus intenciones originales.

Asimismo, los *burócratas a nivel de calle o comunitario*, «trabajadores de la función pública que interactúan directamente con los ciudadanos en el curso de sus empleos», pueden influir en los resultados por medio de la infinidad de decisiones que toman en la interacción con los clientes. Desarrollan conscientemente estrategias de manejo de situación, como limitar la información acerca de sus servicios, pedir a clientes e inspeccionados que esperen, no estar disponibles para los contactos, hacer un sobrado uso de las derivaciones de los clientes a otras autoridades o concentrar la atención en un número limitado de casos fáciles y bien definidos.

Respuesta de los destinatarios. La comprensión, capacidad, buena voluntad, pertenencia organizacional y predicamento de los destinatarios pueden afectar los resultados. Si los clientes no saben ni entienden, la intervención puede no funcionar. Un programa de becas desconocido para los posibles candidatos no contribuirá a alcanzar los resultados para los que fue establecido. La adaptación temporal de una intervención a la situación de decisión de los destinatarios puede explicar los resultados. Si el programa llega en un *momento formativo*, puede producir resultados rápidos, de lo contrario, no. La buena voluntad de los destinatarios individuales también puede jugar un papel preponderante. El programa puede ser acatado de manera más eficiente entre algunos actores debido a los *fanáticos*, personas que están fuertemente comprometidas con la causa y que están dispuestas a dedicar su tiempo y su energía a ella, y que son capaces de hacerlo. Si los fanáticos desaparecen de otras actividades, los resultados deseables pueden desvanecerse.

A menudo, el Estado intenta provocar la acción cuyos destinatarios tienen poco interés en adoptar o que altera el comportamiento que los destinatarios no quieren cambiar. Los regulados también pueden tratar de *capturar* a los reguladores (teoría de la captura regulatoria). Los funcionarios individualmente o toda la agencia pueden ser capturados y comenzar a defender los intereses de aquellos regulados en contra del interés público.

Otra posibilidad es que los destinatarios por propia iniciativa hayan planeado implementar las medidas que el gobierno quiere que implementen, pero que de todos modos participen del programa solicitando y obteniendo apoyo económico. Entonces, los destinatarios son *free-riders* del programa gubernamental. Para sonsacar los efectos programáticos, deben restarse estos supuestos efectos de peso muerto, porque se habrán producido aun sin el programa.

Otras intervenciones gubernamentales, otras agencias del gobierno. Los resultados pueden ser el producto de otras intervenciones y agencias del mismo sector o de otro, ya sea directamente o por medio de complicadas interacciones con las actividades del programa que se valora.

Mercados y otros entornos. El poder de los consumidores en el *mercado* puede influir en la implementación y en los resultados. En las políticas de ambiente, se observó que los productores, debido a una fuerte demanda de los consumidores, se vieron forzados a poner fin a la producción de un *commodity* antes de que las regulaciones gubernamentales determinaran que tenían que hacerlo. La gerencia de una empresa también puede decidir desmantelar aplicaciones industriales que dañan el ambiente antes de tener que hacerlo a causa de una regulación gubernamental o, aunque el gobierno no la obligue, con el objetivo de evitar un anticipado debate público crítico, que puede perjudicar la imagen de la empresa y, por lo tanto, a largo plazo afectar ventas futuras. Por último, los *medios de comunicación*. Una fuerte opinión positiva en los medios a favor de un programa específico fortalecerá psicológicamente a los gestores y a los operadores de las agencias de control, y los hará actuar con mayor firmeza y persistencia para descubrir infracciones.

Utilización

El hecho de que los hallazgos de la evaluación deberían ser útiles y utilizables es un dogma de la comunidad de evaluación. Los precursores de la evaluación en la década de los sesenta sostuvieron una visión *instrumental* de su uso. Los hallazgos de la evaluación se usan en tanto son adoptados por los responsables de tomar las decisiones y pueden influir en su proceso real de toma de decisiones y en la acción

práctica. Esos hallazgos pueden ser sobre medios eficaces para alcanzar determinados fines.

El *uso conceptual* implica que los hallazgos de la evaluación se incorporan al marco de referencia del usuario, pero no se actúan. La evaluación se emplea en el pensamiento relevante para la acción, pero no en la acción. Los políticos y otras partes interesadas tienen discernimientos cognitivos, evaluativos y normativos por medio de la evaluación. Estos pueden contribuir a un escrutinio riguroso de las premisas de la intervención, a un esclarecimiento de sus problemas y a una profunda comprensión de sus méritos y limitaciones. Pero nada de esto se transforma en acción.

También hay una *utilización interactiva* de la evaluación. Interactivo significa que la utilización sucede en procesos de intercambio *deliberativos* y de *diálogo* entre evaluadores y profesionales, así como entre diversos grupos de profesionales. Los diseñadores de políticas públicas están involucrados en procesos de búsqueda que suponen codearse con diferentes actores, donde los evaluadores forman un grupo y su perspectiva constituye únicamente un conjunto entre muchos otros. Los participantes pueden incluir administradores, representantes de los grupos de interés, planificadores, periodistas, clientes, políticos designados, amigos, miembros de partidos y ciudadanos en ejercicio. Todas las partes aprenden. El proceso desde la evaluación hasta la futura decisión no es lineal y unidireccional, sino desorganizado, desordenado e interactivo de una forma que escapa a la representación gráfica.

El uso interactivo involucra la aplicación de información obtenida de informes de evaluación en conjunción con otros datos basados en investigación y antecedentes, como el sentido común, la sabiduría popular, el conocimiento tácito y las propias experiencias directas de los destinatarios. Es probable que este sea el tipo de utilización más interesante a corto, medio y largo plazo.

Las aplicaciones instrumentales, conceptuales e interactivas son utilidades razonables de la evaluación. Sin embargo, existen dos usos adicionales que parecen ser más problemáticos desde un punto de vista normativo y ético.

En el *uso de la legitimación*, se aprovecha la evaluación para justificar posturas establecidas que tienen su fundamento en otras consideraciones, como la ideología política, expectativas electorales, conveniencia de las coaliciones o idiosincrasias personales. La evaluación sirve para fortalecer las posturas positivas o negativas de una persona ante cuestiones o adversarios y aliados políticos. La tarea de la evaluación es proporcionar municiones para las batallas políticas, donde las alianzas ya están formadas y los frentes ya existen.

La *utilización táctica* afirma que la evaluación se solicita para ganar tiempo, evitar la responsabilidad o establecer una arena para elaborar acuerdos. Lo importante es que se pautó una evaluación y que está en marcha, no los posibles hallazgos. Esta observación revela una cuestión crucial. En el uso instrumental, conceptual, interactivo y de legitimación, se utilizan los hallazgos de la evaluación sobre la intervención. En la utilización táctica, se usa el proceso de encargar y de ejecutar la evaluación, no los hallazgos. La evaluación es un concepto producto-proceso; no solo puede usarse el producto, los hallazgos, sino también los procesos previos para hacer la evaluación.

Estrategias para mejorar el uso de los hallazgos de la evaluación

En la Figura 6, se detallan cuatro enfoques para mejorar el uso de los hallazgos de la evaluación.

Figura 6. Estrategias para mejorar el uso de los hallazgos de la evaluación

-
1. **Estrategia orientada a la difusión**
 - a. Método de presentación de informes.
 - b. Método de vinculación.
-
2. **Estrategia orientada a la producción**
-
3. **Estrategia orientada al usuario**
-
4. **Estrategia orientada a la intervención**
-

La *estrategia orientada a la difusión* está interesada en divulgar los hallazgos de la evaluación de la forma más eficaz posible. La evaluación no tiene nada de malo siempre que se realice con las mejores metodologías disponibles. El cuello de botella está en que los destinatarios deben conocer los hallazgos antes de poder usarlos.

Dentro de la estrategia orientada a la difusión, se distinguen dos subenfoques: el método de presentación de informes y el método de vinculación. Al concentrarse en los documentos de la evaluación, en los tratados y en las sesiones informativas orales, el *método de presentación de informes* intenta transmitir del modo más amplio el producto de la evaluación y hacerlo lo más amigable posible para los destinatarios, sin comprometer la metodología ni los hechos.

En la Figura 7 se presenta una serie de estrategias para simplificar la presentación de informes de la evaluación.

Figura 7. La estrategia orientada a la difusión: el método de presentación de informes

Presentación de informes de proceso (primaria y secundaria, antes del informe final)

- Debe localizarse a clientes potenciales de la evaluación en las primeras etapas del proceso de evaluación.
- Debe difundirse la información acerca del inicio de la evaluación.
- Deben difundirse los hallazgos y las recomendaciones a las diversas audiencias de manera rápida y continua antes de que se complete el informe final.

Presentación de informe: diseño y forma del informe final

- Un informe debe contener un resumen muy visible para que los posibles usuarios decidan si continúan leyendo o no.
- Un informe debe contener un resumen ejecutivo, un tanto más largo que el resumen, pero aun así corto y perspicaz.
- Los informes deben mostrar algún hecho sorprendente que haga que las personas se detengan a pensar.
- Los informes deben ser directos y breves.
- Cada informe escrito debe limitarse a un asunto incisivo; si es complejo, los resultados deben presentarse en varios informes breves, no en un tratado exhaustivo.
- Los informes escritos deben confeccionarse en el lenguaje del usuario y no en una jerga diseñada para que las ideas simples sean difíciles de comprender.
- Las consideraciones de los resultados deben estar acompañadas de gráficos, cuadros y otras imágenes.
- Los resultados cruciales deben destacarse estilísticamente mediante el uso de títulos claros, subtítulos y una adecuada organización general del análisis; deben presentarse primero los hallazgos sustantivos y después los métodos; los principales resultados sustantivos deben exponerse con terminología inequívoca antes de las salvedades, y no al revés; es importante que el resumen ejecutivo comience con los principales hallazgos sustantivos.
- La argumentación sobre los métodos debe limitarse al mínimo en el grueso del informe; en cambio, las consideraciones metodológicas deben agregarse como anexos.
- Los informes deben incluir recomendación para la acción.

Presentación de los informes finales: manejo

- Los informes finales deben ser rápidos y oportunos.
 - Los gerentes y otras partes interesadas deben recibir copias impresas del informe final.
 - Los informes finales deben difundirse a través de las bibliotecas, Internet y bases de datos para permitir el uso a largo plazo.
-

Presentación secundaria de los informes finales

- Los resultados finales deben comunicarse en persona.
 - Los evaluadores deben involucrarse en la venta de sus hallazgos finales.
 - Los evaluadores deben estar cerca en caso de que los potenciales usuarios deseen realizar preguntas y hablar.
 - Los evaluadores deben hablar sucintamente y con frecuencia.
 - Los evaluadores deben contar historias —anécdotas de desempeño— para ilustrar las cuestiones.
 - Los evaluadores deben involucrarse en el debate público.
 - Los evaluadores deben difundir los hallazgos finales participando en reuniones, conferencias, seminarios y diálogos.
 - Los evaluadores deben difundir los resultados mucho tiempo después a través de contactos y de redes.
-

Fuente: Adaptado de Vedung, 1997: 281.

El éxito del método de presentación de informes requiere que los evaluadores se comprometan con su utilización, localicen potenciales usuarios, intenten evitar escribir de manera incomprensible, se esfuercen para diseñar ensayos e informes amigables para el usuario y asuman el papel de defensores acérrimos de sus hallazgos.

El propósito del *método de vinculación* es promover la difusión abriendo canales para los profesionales de modo organizado, sistemático y sostenible. Supone el uso permanente de algún agente intermediario entre evaluadores y destinatarios. Las comisiones asesoras son un posible agente intermediario, los líderes de opinión son otro. Involucrar a un especialista en transferencia de información sería otro caso.

Sin embargo, no es suficiente solo una buena presentación de informes. La *estrategia orientada a la producción*, el segundo enfoque importante de uso de las mejoras, plantea que los hallazgos de la evaluación deberían hacerse más pertinentes y de mayor calidad mediante esfuerzos dirigidos al proceso de evaluación.

Los evaluadores deberían llevar adelante una autoevaluación de su propia evaluación, preferentemente el informe final, antes de que quede en manos de los comisionados. La autoevaluación equivale al control de calidad. Otra estrategia es incentivar los comentarios críticos de académicos y colegas familiarizados con el tema pero desapasionados.

La *receptividad a las preocupaciones del usuario* es importante. El evaluador debería prestar atención a las preguntas de los potenciales usuarios, no a las de los académicos. Preferentemente, evaluadores y posibles

destinatarios deberían elaborar juntos las preguntas antes de que queden en manos de los evaluadores para su investigación. O bien, destinatarios y evaluadores deberían trabajar juntos durante todo el proceso de valoración. Las evaluaciones deberían estar dirigidas por la demanda y no por la oferta.

Se recomienda el uso de *variables manipulables* como factores explicativos. Los usuarios solo están atentos a las contingencias que pueden influirse a través de la acción humana o de forma más específica, del propio usuario. Deben evitarse las sugerencias cuyo objetivo es una reestructuración total de la sociedad. Incluso así, la noción subyacente es que las evaluaciones deben ser teóricas para poder ser prácticamente aplicables. No hay nada más práctico que una buena teoría. En este contexto, teoría significa teoría explicativa. Solo cuando sabemos qué factores condicionan un resultado beneficioso o perjudicial, se pueden sacar las conclusiones adecuadas para el futuro.

Algunas personas van más allá y sostienen que los evaluadores deberían centrarse en *variables factibles*. Las modificaciones sugeridas deben ser política y administrativamente aceptables.

Un modelo más trascendental de la estrategia centrada en la producción sostiene que los evaluadores deben consultar a los posibles destinatarios en todas las etapas de la evaluación y de la utilización. Además de la identificación del problema, la planificación para la recolección de datos, la recolección de datos, el procesamiento de datos, la elaboración de conclusiones y la escritura del informe, la cooperación debería extenderse también a la difusión y a la utilización.

El enfoque basado en la consulta de las partes interesadas presenta ventajas. Las posibilidades de brindar el tipo correcto de información a los destinatarios aumentarán, y esto mejorará la probabilidad de que los usuarios se comprometan con los hallazgos, lo que los hará más propensos a usarlos o recomendar a otros que lo hagan.

Otro mérito es que el aprendizaje puede ocurrir en el proceso de evaluación, es decir, mucho antes de la publicación del informe final.

Una desventaja es que los evaluadores se arriesgan a involucrarse en los procesos políticos. Si tuviéramos que elegir entre objetividad y utilidad, es posible que prefiramos esta última. Desde el punto de vista de la investigación académica, resulta importante separar objetividad y verdad de utilidad y uso. Se puede usar conocimiento no válido, y el conocimiento válido puede quedar sin uso. La evaluación no debería politizarse.

Una estrategia indirecta orientada a la producción sería una *metaevaluación* en el sentido de auditar la función de evaluación. Por lo general, este tipo de metaevaluación se incluye en una filosofía más amplia de la gestión de la evaluación que sugiere lo siguiente: en lugar de realmente llevar adelante evaluaciones sustantivas, la administración sénior debería concentrarse en auditar la función de evaluación en oficinas subordinadas. Otra variante es que los niveles más bajos realicen autoevaluaciones y resuman sus hallazgos en un ensayo de evaluación; entonces, la tarea de la gerencia sénior sería evaluar y usar el informe de evaluación.

El tercer enfoque más importante del mejoramiento de la utilización, la *estrategia orientada al usuario*, suscita hacer que los potenciales clientes de la evaluación sean más susceptibles a la utilización. Los usuarios podrían ser educados sobre la evaluación. Para ser utilizadas, las evaluaciones necesitan algún defensor en la organización destinataria. El procedimiento estándar en esta área sería incorporar la evaluación al sistema de gestión de la organización. La evaluación debería institucionalizarse como un asunto interno continuo. La gerencia sénior debe involucrarse enérgicamente.

Una cuarta posibilidad es ajustar el proceso de formación de políticas a las demandas de la investigación de la evaluación. Esta *estrategia orientada a la intervención* podría ejecutarse por medio de un diseño de dos pasos de las políticas públicas: primero, una prueba provisional a pequeña escala acompañada de una rigurosa evaluación. Segundo, la puesta en marcha de la mejor alternativa que se obtuvo mediante la prueba y fue señalada por la evaluación.

Notas

* Título original “*Evaluation research*”, en Guy Peters y John Pierre (comp.), *Handbook of Public Policy*, pp. 397-416, Sage Publications, Londres 2006. Publicación autorizada en inglés © 2006 por Sage Publications. Reproducido bajo licencia de SAGE Publications, London, Los Angeles, New Delhi y Singapore. Todos los derechos reservados.

¹ Fuente: Adaptado de Vedung, 1997, 93f. Vedung, E. (1997). *Public Policy and Program Evaluation*. New Brunswick: Transaction. En español: Vedung, E. (1997). *Evaluación de Políticas Públicas y Programas*. Madrid: Ministerio de Trabajo y Asuntos Sociales, Instituto Nacional de Servicios Sociales.

Referencias

Dale, R. (2004). *Evaluating Development Programmes and Projects*. Segunda edición. Londres: Sage Publications.

Furubo, I., Rist, R. C. y Sandahl, R. (Eds.) (2002). *International Atlas of Evaluation*. New Brunswick, Nueva Jersey y Londres: Transaction Publishers.

Martin, L. L. y Kettner, P. M. (1996). *Measuring the Performance of Human Service Programs*. Londres: Sage Publications.

MEANS Collection (1999). *Evaluation of Socio-economic programmes*. Luxemburgo: Office for Official Publications of the European Communities. ISBN: 92-828-6626-2. Seis volúmenes: (1) Diseño y administración de la evaluación. (2) Selección y uso de indicadores para el monitoreo y la evaluación. (3) Principales técnicas y herramientas de evaluación; (4) Soluciones técnicas para la evaluación con asociación. (5) Evaluación transversal de los impactos en el ambiente, el empleo y otras intervenciones prioritarias. (6) Glosario de trescientos conceptos y términos.

Owen, J. M. y Rogers, P. J. (1999). *Program Evaluation: Forms and Approaches*. Londres: Sage Publications.

Pawson, R. y Tilley, N. (1997). *Realistic Evaluation*. Londres: Sage Publications.

Rossi, P. H., Lipsey, M. W. y Freeman, H. E. (2004). *Evaluation: A Systematic Approach*. Séptima edición. Thousand Oaks, California: Sage.

Stern, E. (Ed.) (2005). *Evaluation Research Methods*. Volúmenes I-IV. Londres: Sage Publications, Sage Benchmarks in Social Research Methods.

Vedung, E. (2001). *Publication Policy and Program Evaluation*. New Brunswick, Nueva Jersey y Londres: Transaction Publishers.

Abordaje de la atribución a través del análisis de contribución: uso sensato de las mediciones de desempeño*

John Mayne

Un elemento significativo de la reforma del sector público en muchas jurisdicciones es el alejamiento de un régimen de administración enfocado en reglas y procedimientos para pasar a un enfoque que le presta mucha atención a los resultados que se buscan con el dinero de los contribuyentes. Los términos «gestionar por resultados», «gestión basada en resultados» y «administración del desempeño» se tornaron comunes en los debates sobre la reforma del sector público (Auditor General de Canadá, 1997; Organización para la Cooperación y el Desarrollo Económicos [OCDE], 1997; Presidente de la Junta del Tesoro, 1997). El objetivo es cambiar la cultura de la administración pública de una centrada en reglas a una enfocada en resultados que les importan a los ciudadanos. Este enfoque se caracteriza por medir el progreso hacia los resultados que se buscan, tiene la flexibilidad de ajustar las operaciones para un mejor cumplimiento de las expectativas e informa sobre los efectos logrados. Algunas jurisdicciones legislaron sobre este enfoque en la administración pública. En muchos casos, se realizaron progresos en esa dirección. Sin embargo, los desafíos de gestionar por resultados han sido y continúan siendo significativos, en especial, la dificultad de medir resultados en el sector público y de establecer conexiones con las actividades del programa de manera rentable. Algunos de estos problemas de medición se discuten a continuación. Existe un problema adicional relacionado que no ha recibido suficiente atención: la necesidad de repensar cómo manejarse con la rendición de cuentas en este nuevo paradigma de la administración.

Rendición de cuentas por los resultados¹

Antes era muy probable que la rendición de cuentas de los procesos implementados, de los insumos utilizados y, quizá, de los productos creados fuera la arena de trabajo de los funcionarios públicos. Este enfoque era consistente con la visión tradicional de la rendición de cuentas: insistir en lo que podía controlarse y buscar culpables cuando las cosas salían mal. Si no se seguía el proceso esperado, se usaban insumos inadecuados o no se entregaba la producción, el responsable podía ser identificado y podían adoptarse las medidas correspondientes, como debería ser en el control de los procesos, los insumos y los productos. Dado este paradigma, los funcionarios públicos a menudo eran reticentes a dar cuenta de los resultados más allá de los productos, en otras palabras, los efectos sobre los que no tienen absoluto control. Para los funcionarios públicos, ha sido mucho más aceptable rendir cuenta por los productos que por los efectos. Y en estos casos, establecer una conexión entre actividades y productos —atribución— no es una cuestión importante: está claro que el programa genera los productos.

En el caso de gestionar por resultados, y en especial por efectos, el grado de control administrativo y el ámbito de influencia de un administrador en los efectos buscados variarán considerablemente en distintas situaciones. En algunos casos, el administrador del programa en cuestión es el principal actor y posee un importante grado de control sobre los efectos. En otros casos, el administrador puede ser solo uno de varios actores que intentan, con los recursos y con las autoridades disponibles, influenciar el logro de los resultados deseados. Una rendición de cuentas eficaz implica que los administradores comprenden estas consideraciones y cuentan con los medios para encargarse de situaciones más complejas.

No lograr los efectos esperados puede deberse a varias razones, de las cuales solo una corresponde a que el administrador «responsable» no hizo un buen trabajo. De hecho, el administrador pudo haber hecho todo lo que podía esperarse, pero aun así los resultados no se alcanzaron debido a circunstancias que exceden su influencia. Con el fin de incentivar y de respaldar la administración por resultados, necesitamos una nueva visión de la rendición de cuentas que reconozca este mundo de la

administración de mayor complejidad (Hatry, 1997). Aquí la atribución es un problema real.

La rendición de cuentas por resultados o efectos² pregunta si usted hizo todo lo posible con las autoridades y los recursos para el logro de los resultados deseados, y si aprendió qué funciona y qué no a partir de la experiencia. Este tipo de rendición por los resultados significa demostrar que usted marcó una diferencia, que contribuyó, mediante sus acciones y esfuerzos, con los resultados alcanzados. Encontrar formas creíbles de demostrarlo es esencial para que tenga éxito el cambio hacia la administración por resultados.

El problema de la atribución

El objetivo de los programas gubernamentales es producir ciertos efectos: más empleos, una sociedad más saludable y mejores condiciones de vida, entre otros. Los programas eficaces son los que marcan una diferencia cumpliendo con este tipo de objetivos —contribuyen con los efectos deseados que los ciudadanos valoran—. Al intentar medir el desempeño de un programa, nos enfrentamos con dos problemas: podemos medir —a menudo, aunque con cierta dificultad— si estos efectos en verdad ocurren o no. La cuestión más compleja es determinar qué contribución hizo exactamente el programa específico al efecto. ¿Cuánto del éxito (o del fracaso) podemos atribuirle al programa? ¿Cuál fue la contribución del programa? ¿Qué influencia tuvo?

A pesar de la dificultad en la medición, la atribución es un problema que no puede ignorarse al evaluar el desempeño de los programas gubernamentales. Sin una respuesta a esta cuestión, poco puede decirse sobre el valor del programa ni pueden darse recomendaciones sobre futuras direcciones. Tal vez, aun sin el programa, los cambios observados en los efectos hubieran ocurrido igual o en un menor nivel o más tarde. En la mayoría de los casos, existen muchos otros factores en juego además del impacto de las actividades del programa. Cuestiones como otras acciones o programas gubernamentales, factores económicos, tendencias sociales y similares pueden tener impacto en los efectos. A los administradores, al gobierno y a los contribuyentes les gustaría conocer la contribución

del programa para evaluar el valor de continuidad en su forma actual. A menos que podamos manejar este problema de medición, la rendición por resultados nunca se afianzará.

La pregunta es: ¿cómo podemos demostrar que un programa marca una diferencia?

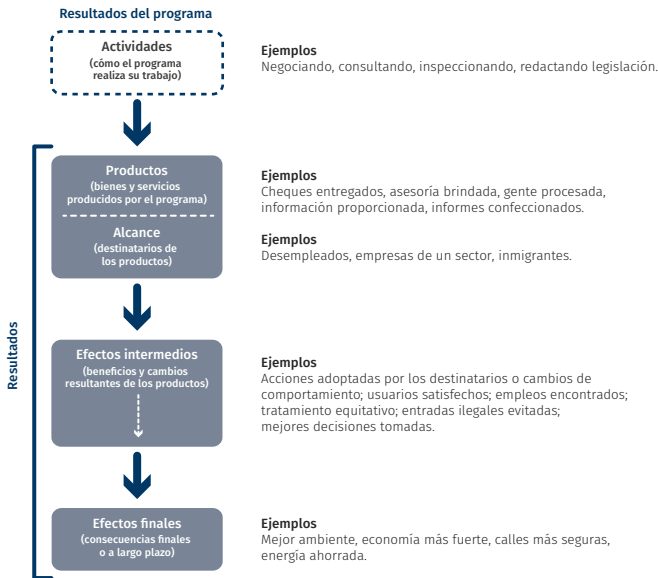
La evaluación de políticas y programas es una disciplina de medición que intenta brindar respuestas al interrogante de la atribución (Freeman y Rossi, 1993; Hudson, Mayne y Thomlinson, 1992; Wholey, Hatry y Newcomer, 1994). Tradicionalmente, se usa algún tipo de comparación controlada para estimar qué sucede con el programa implementado frente a qué ocurriría sin él. Se diseñaron exhaustivos métodos de investigación de las ciencias sociales teniendo en mente este problema de atribución. Y es probable que un estudio de evaluación continúe siendo la mejor manera de abordarlo si uno dispone del tiempo, el dinero y la pericia.

El caso de la medición de desempeño

La medición de desempeño se usa extensamente y cada vez más para medir el desempeño de los programas gubernamentales (Mayne y Zapico-Goni, 1997). A diferencia de la evaluación, que en general realiza mediciones especiales por única vez y análisis exhaustivo de los datos recopilados, la medición de desempeño se caracteriza por medir regular y directamente los aspectos del desempeño de un programa. Los indicadores de desempeño se usan para monitorear desempeño y como información de retroalimentación para los administradores y para el personal. Pueden conformar la base de los informes sobre logros del programa.

A menudo, la medición de desempeño se dirige al primer nivel de impacto de un programa, es decir, a medir los productos específicos (bienes y servicios) que el personal del programa provee (ver Figura 1).

Figura 1. Tipos de resultados de programas



En estos casos, es improbable que la cuestión de la atribución sea un problema porque existe una conexión directa evidente entre lo que el personal hace y sus productos inmediatos. No obstante, cada vez más las organizaciones tratan de medir o rastrear los impactos subsecuentes de estos productos y servicios, los resultados intermedios o incluso los resultados finales que intentan lograr. El asunto de la atribución emerge con rapidez. Ante la ausencia de un riguroso estudio de evaluación, ¿qué puede hacerse?

Es posible estructurar un sistema de medición de desempeño para intentar obtener directamente una medición de la atribución. Se podrían construir series de datos temporales modificando el programa a lo largo del tiempo y haciendo un seguimiento de los cambios resultantes. O bien, además de medir los impactos en quienes reciben el programa, también podrían medirse los cambios sucedidos comparando con un grupo similar que no recibe el programa. Estos enfoques devienen en evaluaciones *de facto* usando algún tipo de diseño cuasiexperimental.

Aunque posible, esta estrategia de medición cuidadosamente construida y generalmente costosa no se asocia con la mayoría de los enfoques de medición de desempeño. Ante la falta de un estudio de evaluación, ¿qué puede hacerse en caso de una medición de desempeño o de un sistema de monitoreo «normal» o típico para la cuestión de la atribución? Este es el interrogante que se aborda en este trabajo.

Reconocer los límites de la medición

Primero, debemos reconocer que, por lo general, no es posible determinar el alcance por el que un programa gubernamental contribuye a un efecto particular, aun con un estudio de evaluación cuidadosamente diseñado. Podríamos ser capaces de brindar considerable evidencia relacionada con los impactos de un programa y de incrementar significativamente nuestra comprensión sobre el modo en el que un programa influye en cierto efecto, pero no garantizado al cien por cien en la mayoría de los casos de cualquier complejidad. En cambio, debemos hablar de reducir nuestra incertidumbre sobre la contribución del programa. A partir de un estado de no saber nada sobre el modo en el que un programa influye en un efecto deseado, podríamos concluir con razonable confianza que el programa está teniendo un impacto atribuible, que marca una diferencia. También podríamos ser capaces de estimar razonablemente la magnitud del impacto.

Así, es posible que debamos repensar la utilidad de la medición. La medición en el sector público no se trata tanto de precisión, sino más bien de aumentar la comprensión y el conocimiento, de incrementar lo que sabemos sobre lo que funciona en un área y así reducir la incertidumbre (Auditor General de Canadá, 1996: 21). Esta visión de la medición implica que casi siempre podemos medir cosas y, en particular, la contribución que hace un programa. En otras palabras, podemos recopilar información y datos adicionales que aumentarán nuestra comprensión acerca de un programa y sus impactos, aun si no podemos «demostrar» ciertas cuestiones de manera absoluta. Debemos incluir herramientas de medición cualitativas y más blandas en nuestro concepto de medición para el sector público.

Los límites de la medición se refieren a que la necesidad de aceptar cierta incertidumbre sobre las mediciones de desempeño que probablemente tengamos a disposición en muchos casos. Para conocer con un alto grado de certeza cuál es la contribución de un programa, se requiere una evaluación bien diseñada. Lo que abordo en este trabajo se aplica a casos en los que una persona podría conformarse con menor certeza —o así se le exige— en los que el objetivo de la medición es adquirir cierto discernimiento y desarrollar cierta certeza de que el programa tiene impacto. Sugiero que este es o debería ser el objetivo de la medición de desempeño. Una buena estrategia global de medición incluiría tanto una medición permanente de desempeño como una evaluación periódica.

Dos usos de la medición de desempeño: comprender e informar

Es posible distinguir dos usos de la información sobre la medición de desempeño. Primero, puede emplearse para comprender mejor qué contribución hace exactamente un programa. Esta es la perspectiva de la administración, en la que la medición sirve para saber más sobre si el programa marca una diferencia y de qué modo lo hace; se busca conocimiento. Se quiere determinar si el programa es la herramienta adecuada en materia de políticas para alcanzar el resultado deseado. Aquí la pregunta es cómo usar la medición de desempeño como una herramienta de investigación.

Segundo, sirve para explicar o demostrar el desempeño alcanzado por un programa. En muchas jurisdicciones, existe un mayor énfasis en informar a los funcionarios elegidos y al público qué se logró con el dinero recaudado y con los recursos utilizados. Es frecuente que las medidas de desempeño formen las bases de este reporte. Aquí el interrogante es: ¿cómo se puede usar la información sobre la medición de desempeño para informar con credibilidad acerca de lo que se logró? En especial, ¿cómo puede usarse para informar de la mejor manera sobre la contribución que realiza un programa?

Debemos tener en cuenta estos dos usos mientras consideramos cómo manejarnos con la atribución usando medidas de desempeño.

Enfoques para la atribución: análisis de contribución

Tanto para la comprensión como para la presentación de informes se necesita un análisis específico que brinde información de la contribución de un programa a los efectos que intenta influenciar. Junto con los comentarios anteriores sobre la naturaleza de la medición en el sector público, la tarea a realizar podría describirse mejor si decimos que, para el caso de la presentación de informes, se trata de representar una imagen creíble sobre la atribución de un programa. Para la comprensión, la tarea es recopilar, a partir de las medidas de desempeño, el mayor discernimiento posible de cuán bien funcionan las operaciones de un programa. En ambos casos, la herramienta clave para el análisis es la cadena de resultados, que ilustra lo que sucede como resultado de las actividades que se realizan y de los productos que se crean.

Muy a menudo, la medición de desempeño, y en particular la presentación de informes por medio de sistemas de medición de desempeño, ignoran el problema de atribución. El desempeño medido es atribuido directamente al programa o atribuido por implicancia, a través de la falta de discusión o de análisis de otros factores en juego. Para cualquiera, aun con escaso conocimiento del programa y de su entorno, esta clase de información sobre el desempeño tendrá poca credibilidad. Para los administradores, no brinda información con valor agregado. En la mayoría de los casos, puede anticiparse cualquier número de factores para explicar el efecto observado que no sea el propio programa. Cuanto más obvios son estos factores, menos creíble la información sobre el desempeño. Discutir otros factores también puede proveer discernimiento del propio programa, de cómo funcionó y de sus impactos.

Reconocer el problema

Es necesario reconocer que, además del programa, hay otros factores en juego y que, por lo tanto, no queda claro inmediatamente qué efecto tuvo o tiene el programa en producir el resultado en cuestión. Los administradores deben ser realistas sobre los efectos que intentan influenciar si desean obtener nueva comprensión acerca de si sus actividades marcan una diferencia y cómo lo hacen. Para la presentación de informes, reconocer

otros factores en juego es más honesto y, en consecuencia, más creíble que fingir que no existen. Como veremos más adelante, se puede hacer más, pero reconocer otros factores en juego mientras todavía se cree que el programa hace una contribución es un primer paso crítico.

Sugerimos una serie de pasos que se detallan en la Figura 2 y que pueden usarse para abordar la atribución a través de la medición de desempeño. Colectivamente, estos son elementos de un *análisis de contribución*. Este análisis intenta explorar y, tal vez, demostrar lo que Hendricks (1996) denomina «asociación plausible»: si «una persona razonable, al saber qué ocurrió en el programa y que los efectos deseados en verdad sucedieron, está de acuerdo con que el programa contribuyó con esos efectos».

Figura 2. Análisis de contribución

Paso 1: Desarrollar la cadena de resultados. Desarrollar el modelo teórico/modelo lógico/cadena de resultados del programa que describan cómo se supone que trabaja. Identificar los principales factores externos en juego que podrían explicar los efectos observados. Este modelo teórico del programa debería conducir a una asociación plausible entre las actividades del programa y los efectos buscados. Algunas conexiones en la cadena de resultados se comprenderán o aceptarán sin inconvenientes. Otras generarán confusión o estarán sujetas a otras explicaciones que no sean que el programa fue la «causa». De esta forma, se reconoce que la atribución constituye, en efecto, un problema.

Paso 2: Evaluar la evidencia existente sobre los resultados. La cadena de resultados debería proporcionar una buena idea de qué resultados deseados (productos, resultados intermedios y finales) podrían medirse. ¿Qué evidencia (información obtenida de medidas de desempeño y evaluaciones) se encuentra actualmente disponible sobre la ocurrencia de estos diversos resultados? También se deben evaluar las conexiones en la cadena de resultados. ¿Cuáles son fuertes (mucha evidencia disponible, lógica fuerte o amplia aceptación) y cuáles son débiles (poca evidencia disponible, lógica débil o escaso acuerdo entre las partes interesadas)?

Paso 3: Evaluar las explicaciones alternativas. Por definición, los efectos se ven influenciados no solo por la acción del programa, sino también por factores externos —otros programas, factores sociales y económicos—. Además de evaluar la evidencia existente sobre resultados, es necesario considerar explícitamente el grado de influencia de estos factores externos. Cierta evidencia o argumentos lógicos podrían sugerir que algunos tienen poca influencia, mientras que otros pueden afectar significativamente los resultados deseados.

Paso 4: Armar la historia de desempeño. Con esta información, se podrá proponer una historia de desempeño y explicar por qué es razonable asumir que las acciones del programa contribuyeron (en cierto modo que se puede abordar o describir) a los efectos observados. ¿Cuán creíble es la historia? ¿Están las personas sensatas de acuerdo con la historia? ¿Valida el patrón de resultados observado la cadena de resultados? ¿Cuáles son los principales puntos débiles? Siempre habrá debilidades, estas señalan las áreas en las que sería útil obtener datos o información adicionales. Si no es posible obtener evidencia adicional (o no lo es por el momento), entonces, esto es lo máximo que puede decirse sobre la medida en la que el programa marcó una diferencia.

Paso 5: Buscar evidencia adicional. Para mejorar la historia de desempeño, se necesita evidencia adicional, que podría ser información tanto del alcance de la ocurrencia de resultados específicos en la cadena de resultados como del peso de ciertas conexiones dentro de esta. En el presente trabajo, se detallan algunas técnicas de refuerzo que se recomienda adoptar.

Paso 6: Revisar y reforzar la historia de desempeño. Con la evidencia nueva, debería poder elaborar una historia más confiable, una con la que es probable que cualquier persona sensata estuviera de acuerdo. No será una historia infalible, pero tendrá más peso y credibilidad.

Paso 1: Desarrollar la cadena de resultados

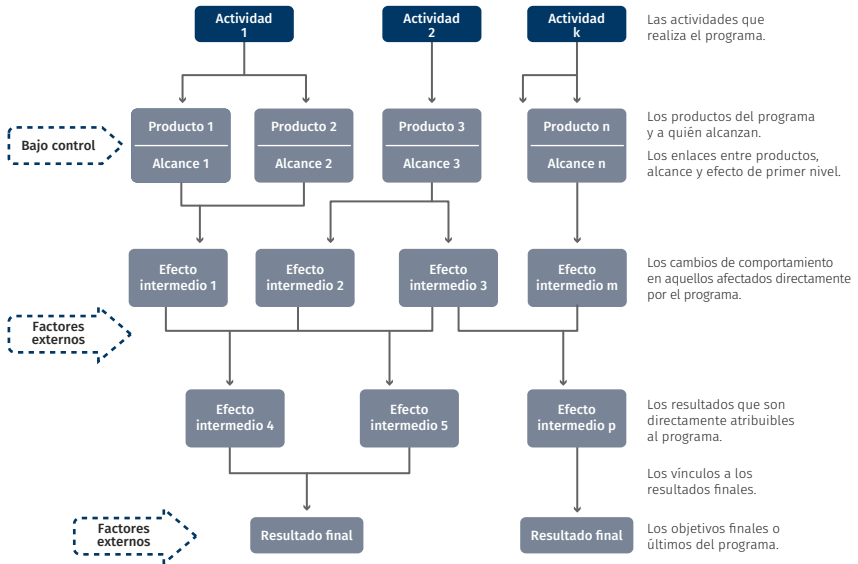
Existe cierto razonamiento lógico detrás del programa que explica qué debería lograrse y cómo. Esta lógica o modelo teórico podría ser bastante convincente o estar bien establecida sobre la base de la experiencia previa. Al desarrollar el caso lógico, podemos ver qué se supone o se cree que sucede. Construir y presentar este modelo teórico del programa es un componente estándar de la planificación para un estudio de evaluación (Wholey, 1983), en la que, por lo general, se usa un cuadro lógico (Julian, Jones y Devo, 1995).

Últimamente, el poder de este enfoque es cada vez más usado en la medición de desempeño, en el que términos como «cuadros de secuencia de efectos», «cadenas de resultados» y «árbol indicador visible» (Meekings, 1995) se usan para describir la misma herramienta de diagnóstico. Además de tratar la cuestión de la atribución, estas herramientas demuestran ser inestimables para el diseño y la implementación de sistemas de medición de desempeño. Asimismo, al forzar a los diseñadores a ser claros sobre los problemas que los programas fueron diseñados para abordar y sobre cómo lo hacen, los modelos lógicos o las cadenas de resultados fomentan un diseño de programa más preciso.

Una cadena de resultados o cuadro lógico trata de mostrar en una página cómo trabaja el programa —es decir, cómo se cree que los diversos productos provocan un número de resultados que producirán los efectos finales deseados—. Las cadenas de resultados también pueden discutir los impactos no deseados que podrían suceder y que deben vigilarse, así como los factores externos clave que influyen en los efectos.

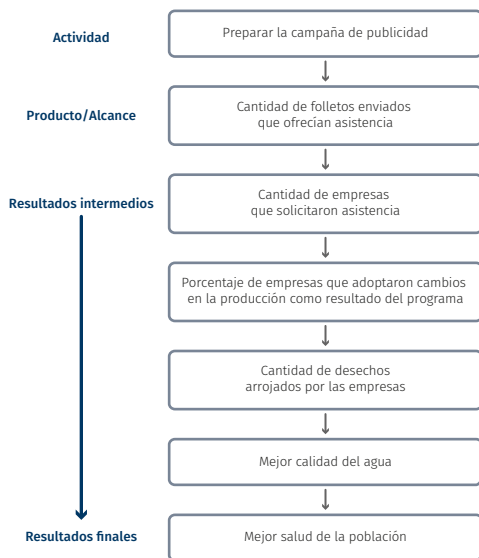
La Figura 3 muestra, de forma genérica, la apariencia de un cuadro lógico detallado; existen muchas maneras de presentarlo. Un cuadro lógico ilustra las conexiones entre productos específicos, resultados intermedios específicos y resultados finales específicos. En otros casos, puede ser suficiente una imagen menos complicada de la lógica del programa.

Figura 3. Cuadro lógico o cadena de resultados del programa



La Figura 4 ilustra este caso para un programa ambiental. Los cuadros lógicos explícitamente incluyen la idea de alcance —a quién se espera que el programa alcance (Montague, 1998)— y los resultados intermedios. Esto es así porque, en general, en estos niveles, los indicadores de desempeño pueden hacer un buen trabajo de medición; estos son los niveles de la cadena de resultados sobre los cuales el programa, a menudo, tiene mayor control. Asimismo, la evidencia de que los efectos deseados más inmediatos ocurrieron es un paso crítico para exponer el resto de la historia de desempeño. Así puede mostrarse que el programa tuvo cierto efecto.

Figura 4. Cadena de resultados



Fuente: Banco Mundial (1996).

Para los administradores del programa, diseñar y usar un cuadro lógico tiene diversos beneficios, como crear consenso sobre lo que intenta lograr el programa; desarrollar comprensión de cómo funciona identificando a los clientes del programa con claridad; buscar y obtener acuerdo sobre qué resultados precisos se desean —expectativas de desempeño—; identificar medidas de desempeño clave. En particular, estamos interesados en los beneficios adicionales de identificar:

- la relación causa-efecto implícita en el modelo teórico del programa;
- los factores externos en juego;
- las áreas en las que la comprensión sobre el impacto del programa es débil.

Algunas de las «conexiones» entre resultados en el cuadro lógico son bien conocidas y se establecieron en prácticas anteriores. Hay menos probabilidades de estar en desacuerdo sobre su papel para lograr los efectos deseados. En cambio, otras conexiones pueden no ser tan bien aceptadas

y son las que sugieren dónde buscar más provechosamente mayor evidencia (es decir, medidas adicionales de desempeño). Toda evidencia adicional que pueda recopilarse para confirmar dichas conexiones sumará comprensión sobre cómo funciona el programa y reforzará el argumento de que hace una contribución. Asimismo, si se identifican factores externos significativos como posibles generadores de impacto en el efecto deseado, la evidencia para refutar o determinar el grado de influencia de ese efecto será útil para abordar la cuestión de la atribución³.

De este modo, los administradores pueden emplear una herramienta de diagnóstico como la cadena de resultados para comprender mejor cómo ellos y otros creen que funciona el programa. Pueden diseñar operaciones del programa que se ajusten a estas expectativas. Mediante la presentación y discusión de la lógica detrás del programa al informar desempeño, se expone con exactitud qué se mide y cuáles son los principales supuestos en materia de contribución del programa. En consecuencia, se identifican las debilidades de los supuestos del programa y se sugiere dónde se necesita más evidencia. Como mínimo, esta clase de presentación de informes permite conocer qué desafíos a la credibilidad de las medidas de desempeño pueden ponerse de relieve.

Paso 2: Evaluar la evidencia existente sobre los resultados

La cadena de resultados debería apuntar a cuáles resultados deseados (productos, resultados intermedios y finales) podrían medirse y a cuáles deberían medirse. Debe observarse qué evidencia (información obtenida de medidas de desempeño y evaluaciones) se encuentra actualmente disponible sobre la ocurrencia de estos diversos resultados. Las conexiones de la cadena de resultados también deben evaluarse: ¿cuáles están bien respaldadas (mucha evidencia disponible, lógica fuerte o amplia aceptación) y cuáles son débiles (poca evidencia disponible, lógica débil o escaso acuerdo entre las partes interesadas)? Esta evaluación de la información disponible sobre la cadena de resultados brindará una buena idea de dónde la historia de desempeño es fuerte y, más importante aún, dónde es débil y puede necesitar datos o información adicionales.

Paso 3: Investigar y discutir explicaciones alternativas plausibles

El problema de atribución surge cuando se cree o se intenta determinar que ciertos efectos son consecuencia del programa y hay explicaciones alternativas plausibles. En otras palabras, los escépticos a que en verdad lo que sumó haya sido la contribución del programa señalarán otras razones para explicar el efecto observado —por ejemplo, otros programas gubernamentales relacionados, tendencias económicas o sociales o comportamiento no afectado por el programa—.

En general, tratar explícitamente con estas explicaciones alternativas es la mejor forma de reforzar un argumento a favor del impacto del programa y la contribución que tiene. Esto conlleva a:

- identificar explicaciones alternativas más probables y la evidencia asociada a ellas;
- presentar cualquier contraevidencia o argumento y, cuando correspondiera, descartar estas explicaciones alternativas;
- ofrecer cualquier evidencia de que el programa es la explicación más probable.

Recordemos que aquí el asunto no es una situación 0-1. Es posible que, en la mayoría de los casos, haya un número de factores que, de hecho, influyen en el efecto en cuestión. Ningún factor, incluido el programa, «causa» el efecto. Por el contrario, el efecto observado es el resultado de un número de factores que contribuyen. Queremos comprender mejor si el programa es un factor significativo en la ocurrencia del efecto y si es probable que el efecto no hubiera sucedido o hubiera ocurrido de modo significativamente diferente sin el programa en marcha.

Con una o más explicaciones alternativas plausibles para la ocurrencia del efecto y con poca evidencia que contrarreste estas explicaciones, tal vez deba concluirse que en verdad no se sabe cuál fue la contribución del programa y deba sugerirse la necesidad de una evaluación o de más evidencia (ver más adelante).

El tipo de evidencia para contrarrestar argumentos alternativos al programa depende de este último y de su situación. Dos tipos genéricos están disponibles: en primer lugar, un argumento lógico. Nos podríamos referir al modelo teórico detrás del programa —la cadena de resultados— y al tipo de teoría necesaria para respaldar las afirmaciones de hipótesis rivales. Apoyar las explicaciones alternativas puede implicar supuestos más improbables que los asociados con el programa. En segundo lugar, puede presentarse evidencia real sobre las explicaciones alternativas para que sea respaldada, como se discute más adelante.

Abordar así el problema de la atribución demuestra que usted:

- está consciente de la complejidad de la situación;
- reconoce y comprende los otros factores en juego.
- concluye (si asumimos que esto sucede), sin embargo, que la explicación más probable para el efecto observado es que el programa hizo una contribución significativa.

Otros pueden tratar de demostrar que algún factor distinto del programa fue el principal en la cadena de eventos que condujeron al efecto, pero usted presenta la mejor evidencia para su caso. A menos que se discutan con franqueza explicaciones alternativas, su argumento sobre la eficacia del programa puede desafiarse efectivamente solo señalando la existencia de explicaciones alternativas.

Paso 4: El armado de la historia de desempeño

Con la información recopilada, incluida la vinculada con las explicaciones alternativas, podrá proponerse una historia de desempeño de por qué es razonable asumir que las acciones del programa contribuyeron a los efectos observados. Y, tal vez, quiera probar y caracterizar la naturaleza y el alcance de la contribución afirmada. ¿Cuán creíble es la historia? ¿Están las personas sensatas de acuerdo con la historia? ¿Valida el patrón de resultados observado la cadena de resultados? ¿Cuáles son los principales puntos débiles? Siempre habrá debilidades. Estas señalan las áreas en las que sería útil obtener datos o información adicionales, que dependen de la

situación en particular. En la actualidad, lo que se tiene es lo que se tiene. En muchos casos, sería mejor concluir que no hay mucha certeza sobre el alcance de la contribución del programa en este momento.

Paso 5: Buscar evidencia adicional

Una vez identificadas las debilidades en la historia de desempeño actual, existen muchos modos de entender mejor la cadena de resultados mediante la recopilación de evidencia adicional. Construir una historia de desempeño debería considerarse un proyecto de múltiples años —un viaje— que se agrega y se modifica a medida que se gana conocimiento por medio del monitoreo y que las circunstancias externas cambian. A continuación, se detallan varios enfoques para fortalecer las historias de desempeño.

Paso 6: Revisar y fortalecer la historia de desempeño

A medida que se recopila y se evalúa nueva evidencia y la cadena de resultados se completa con más información, surgirá una historia más sólida. Por supuesto, más evidencia puede no respaldar las historias de desempeño previas. La contribución del programa puede no ser tan fuerte como se esperaba originalmente, con la evidencia señalando quizá la necesidad de alterar sus actividades.

Técnicas para fortalecer la historia de desempeño

En muchos casos, si una historia de desempeño se construye solo con información disponible, existirán numerosas debilidades en la evidencia y en la lógica. La primera iteración de un análisis de contribución señalará estas debilidades, como se detalla más arriba. Un aspecto clave es encontrar modos de fortalecer la evidencia sobre la contribución del programa. La Figura 5 detalla varias técnicas de fortalecimiento que pueden usarse y que se discuten más adelante.

Figura 5. Técnicas para fortalecer un análisis de contribución

Cualquiera de estas técnicas o todas ellas podrían adoptarse para fortalecer la historia del desempeño. La selección depende de la situación específica.

Refinar la cadena de resultados. A menudo, es posible desarrollar una cadena de resultados más detallada, que muestre cómo un resultado intermedio lleva a otro en un nivel micro. Por ejemplo, el objetivo de muchos programas es cambiar el comportamiento del grupo objetivo por medio de la comunicación. Entonces, los pasos intermedios son asegurarse de que el grupo está consciente de los mensajes comunicados, los entiende y los considera razonables. Cada uno de estos efectos puede medirse.

Obtener más resultados y/o evidencia de conexiones. Aunque es extraño recopilar datos en todos los pasos de una cadena de resultados, dado que en general esto no es rentable, sabemos que podrían recopilarse otros datos. En el caso en el que la cadena de resultados ya está razonablemente detallada, pueden realizarse esfuerzos para medir más los pasos intermedios de la cadena. Si se desarrolló una cadena de resultados más refinada, datos adicionales más refinados podrían recopilarse para comprender mejor los aspectos clave de la cadena de resultados.

Relevar la opinión de otros con conocimiento. Es probable que varias personas que estén informadas, tengan una visión sobre la medida en la que el programa hizo una diferencia, tanto global como en algún punto intermedio de la cadena de resultado. Entre otras, destinatarios del programa, asociados involucrados en la prestación o expertos externos en el área.

Realizar el seguimiento de las variaciones del programa y de sus impactos. Cuando sea posible identificar las diferencias del programa, como a lo largo del tiempo, entre lugares de prestación y entre diferentes grupos objetivo, recopilar datos sobre las correspondientes diferencias en los patrones de resultados puede proporcionar valiosa evidencia que respalde la historia de desempeño completa.

Emprender estudios de caso. Analizar uno o algunos estudios de casos específicos dentro del programa puede brindar evidencia confirmatoria que demuestre que, en estos casos, el programa parece marcar (o marcar) una diferencia. Por sí mismo, uno o dos estudios de casos pueden no ser demasiado creíbles, pero como parte de un paquete de evidencia en el contexto de una cadena de resultados bien desarrollada, apoyan la razonabilidad del modelo teórico del programa.

Identificar otra investigación o evaluación relevante. En algunas áreas del programa, puede haber conclusiones de investigaciones o evaluaciones que respaldan la evidencia para la cadena de resultados o partes de ellas.

Usar múltiples líneas de evidencia. Combinar la evidencia de los diversos enfoques antes mencionados construye un caso más fuerte y convincente que cualquier argumentación.

Realizar una evaluación. Conducir una evaluación del programa puede ser el único modo de recopilar más información creíble sobre la contribución de un programa. De hecho, la medición efectiva de los resultados debería considerarse como una iniciativa de múltiples años que incluye tanto las medidas de desempeño continuo como la evaluación periódica. La evaluación del programa debería ser parte de una buena *estrategia de medición*.

Refinar la cadena de resultados

Quizá necesitemos comprender mejor el programa en cuestión para ahondar en lo que ocurre como consecuencia de sus actividades. Un enfoque es una cadena de resultados más detallada cuyo punto central está en los cambios específicos de comportamiento que derivan de las implicancias del programa que podemos observar para los que son «alcanzados» por el programa. Para provocar un efecto, los programas tienen que modificar el comportamiento de las personas⁴. Los productos del programa deben tener como objetivo influenciar a los clientes o al público objetivo —el elemento de alcance de la Figura 1— para actuar de distintas maneras, de modo que puedan ocurrir los efectos anticipados.

A menudo, las cadenas de resultados se centran en la secuencia de eventos que se espera que ocurra y así pueden estar a un nivel demasiado agregado para detectar cambios de comportamiento específicos que deben suceder como prerrequisitos de cada uno de los eventos. Al intentar identificar y después documentar los cambios en actitudes, conocimiento, percepciones y decisiones de los grupos objetivo que, lógicamente, se vinculan con los efectos observados, en general, podemos obtener una adecuada comprensión del impacto real que el programa está teniendo. Si podemos observar estos cambios a corto plazo, puede mejorarse el caso lógico para la atribución del programa. Asimismo, por lo general, ciertos resultados intermedios pueden medirse con mayor facilidad. Por consiguiente, puede resultar útil establecer expectativas y objetivos de desempeño en este nivel, con un grado razonable de control (Oficina de Rendición de Cuentas del Gobierno de los Estados Unidos [GAO, por su sigla en inglés], 1998).

Un modelo lógico así nos acerca bastante a enfoques de evaluación basada en la teoría (Chen, 1990, 1994; Weiss, 1997), en los que se desarrollan modelos lógicos más detallados del programa para reconstruir la teoría; se evalúa/prueba la credibilidad de los micropasos de la teoría (conexiones de la lógica/historia del programa) y se desarrollan y/o confirman los resultados alcanzados por el programa. Conectar enfoques sobre la medición de desempeño con evaluación basada en la teoría es un área que necesita más investigación.

Los administradores que tratan tanto de comprender mejor los efectos de sus programas como de informar sobre el desempeño pueden beneficiarse ampliando el análisis de los cuadros lógicos para incluir la consideración de cambios específicos de comportamiento esperados como resultado del programa.

Existe un modo adicional en el que un modelo lógico podría refinarse provechosamente. En general, una buena cadena de resultados para un programa ilustra los múltiples aspectos del desempeño que podrían medirse e informarse. Se necesita un cuidado considerable para seleccionar los indicadores de desempeño. Es importante usar los que mejor discriminan los efectos en cuestión o se centran en ellos, aunque a menudo los indicadores usados se relacionan solo en sentido amplio con

las circunstancias de los clientes del programa, con la economía y con la sociedad en su conjunto. Si se le repara en el modo en el que funciona el programa (a partir del análisis de la cadena de resultados), en general, se pueden refinar los indicadores para que se centren más cuidadosamente en los beneficios específicos esperados que el programa debe alcanzar. En particular, se puede tratar y «refinar el denominador» del indicador⁵.

A modo de ejemplo, muchos indicadores son ratios en los que el denominador califica al numerador. Consideremos un programa diseñado para reducir los accidentes aéreos por medio de la inspección de las condiciones de vuelo de un avión. Un indicador podría ser el número de accidentes aéreos por milla aérea volada. Un mejor indicador sería el número de accidentes aéreos por fallas estructurales por milla aérea volada. Pero las fallas estructurales pueden ocurrir a pesar de las inspecciones. Por lo tanto, puede ser aún mejor usar dos indicadores: número de accidentes aéreos por milla aérea volada debido a fallas estructurales en aviones inspeccionados y número de accidentes aéreos por milla aérea volada debido a fallas estructurales en aviones no inspeccionados. Comparando las fallas estructurales en aviones inspeccionados y no inspeccionados, puede estimarse qué tipo de inspección reduce los problemas que los que esa inspección fue diseñada. Cuestiones sobre la atribución persisten, aunque los indicadores más refinados reducen el problema y mejoran la posibilidad de proporcionar información útil acerca de la contribución del programa.

Obtener más resultados y/o evidencia de conexiones

La medición de desempeño se ocupa de reunir evidencia del desempeño del programa. Nuestro análisis de contribución tendrá elementos identificados de la cadena de resultados, en la que la evidencia es débil o nula. Podemos reforzar la cadena de resultados con más y mejor evidencia.

Como se sugirió antes, podría obtenerse evidencia sobre las explicaciones alternativas del efecto observado. Esto significa recopilar datos, como la información contextual e histórica sobre la plausibilidad de las explicaciones alternativas. Los datos pueden ser parte del sistema habitual de medición de desempeño, pero probablemente sean recopilados oportunamente cuando se emprende el análisis de la contribución del programa. La recopilación

de datos podría implicar una revisión de la literatura relevante, encuestas, seguimiento de factores externos relevantes, trabajos de campo o grupos focales. Cuanto más fuerte pueda hacerse el caso, más fuerte será la conclusión sobre la contribución del programa.

Por lo general, dos fuentes de datos útiles son subestimadas. Con frecuencia, considerables datos de los *archivos del programa* están disponibles, algunos de los cuales servirían para brindar información de la contribución del programa. Este tipo de datos existentes, que probablemente fueron recolectados para otros fines, pueden contener información valiosa, en especial si se usa junto con nuevos datos. En otros casos, pueden estar disponibles útiles *análisis secundarios* —estudios que otros hicieron en el área del programa que podrían clarificar cuestiones de la medición y de la atribución—. Incluso en otros casos, puede haberse realizado *metanálisis* —análisis que sintetizan un número de estudios en un área determinada—.

Relevar la opinión de otros con conocimiento

Como otra fuente de evidencia, podría reunirse información sobre la contribución del programa directamente, casi siempre a través de la *opinión de expertos*. En muchas situaciones, hay personas externas al programa que son consideradas conocedoras del área, de los impactos o del entorno en el que funciona el programa. Una encuesta estructurada puede proporcionar cierta evidencia, aunque subjetiva en su naturaleza, de la medida en la que el programa influye en un efecto. En general, se encuesta a esas personas para averiguar más información del programa, en cuyo caso agregar preguntas sobre la atribución no es muy costoso. Un grupo focal de expertos puede ser otro enfoque y permitiría cierto sondeo de por qué se sostienen algunas opiniones. Ante la falta de otros datos más costosos, este enfoque puede ser un modo relativamente económico de aumentar la confianza en la influencia del programa⁶.

Realizar un seguimiento de las variaciones del programa y de su impacto

En los casos en los que las actividades del programa variaron con el tiempo, demostrar que los efectos cambiaron de manera consistente con las variaciones de las actividades puede fortalecer el argumento de que,

de hecho, las actividades marcaron una diferencia. En un ejemplo simple, si se observó un efecto esperado después (y no antes) de que la actividad del programa se pusiera en marcha, esto sugiere que el programa tiene impacto. En un caso más complejo, si el efecto mejora en algunos lugares (o momentos) en los que el programa se implementó, pero no en otros (como un programa nacional que funciona en muchas zonas), el argumento de que marca una diferencia es aún más fuerte. Hendricks (1996) identifica algunos de dichos casos, en los que al realizar un seguimiento de las medidas de desempeño, podríamos demostrar que los efectos:

- aparecieron en un momento apropiado después de que comenzaron nuestros esfuerzos;
- surgieron en diferentes lugares o con distintas personas;
- se desvanecieron cuando nuestros esfuerzos cesaron;
- son solo aquellos que deberíamos haber afectado;
- aparecieron solo dónde y cuándo estuvimos activos;
- más grandes aparecieron donde hicimos el mayor esfuerzo.

En algunas áreas de la programación, como los impactos de las actividades de investigación, es probable que se observe una demora significativa antes de que los efectos deseados ocurran, y la foto de la atribución retratada por medio del seguimiento del desempeño a lo largo del tiempo no será tan evidente. En estos casos, debe realizarse un seguimiento de los efectos en el tiempo para ver si los efectos deseados ocurrieron, aunque demostrar o comprender la atribución es más difícil. Es necesario usar algunos de los otros enfoques descritos en este artículo.

El seguimiento de las variaciones del programa es un enfoque que puede acercarse a los métodos tradicionales de evaluación para analizar la atribución.

Emprender estudios de caso

A menudo, se puede hacer uso de la *evidencia del estudio de casos* sobre los efectos de un programa. En los programas en los que hay casos, proyectos

o eventos específicos, la evidencia para la atribución en uno o dos de ellos puede ser bastante convincente, revelar la verdadera naturaleza del programa y también demostrar, al menos en estos casos, que se puede estar relativamente seguro del impacto de las actividades del programa. Además, los estudios de casos también pueden ilustrar si la lógica del programa es o no sensata y razonable. Este tipo de evidencia puede ser bastante convincente, pero se deben tomar los recaudos correspondientes, en especial cuando la evidencia es anecdótica. El estudio de casos y la evidencia anecdótica son más convincentes cuando se ilustra un caso concreto para complementar otra evidencia recopilada. No obstante, por sí mismo, puede ser bastante engañoso, dado que simplemente puede ser uno de los pocos casos que funcionó entre una vasta mayoría que no funcionó, como la EUA GAO (1996) encontrada en una revisión de «historias de éxito» del Departamento de Medio Ambiente de los Estados Unidos. Asimismo, los lectores se sienten tentados a generalizar a partir de la evidencia anecdótica, una práctica contra la que debería advertirse. Sin embargo, si se aclaran el contexto y las limitaciones, los estudios de casos individuales son útiles.

Usar múltiples líneas de evidencia

Hemos debatido una serie de formas de ocuparse del problema de la atribución. Sugerimos que cuantas más maneras se usen en un caso cualquiera, mayor información definitiva tendremos sobre la atribución. Este es el argumento de las «múltiples líneas de evidencia». Si ninguna evidencia puede ser muy convincente, un gran conjunto de evidencias diferentes y complementarias puede serlo. De este modo, al tratar de reducir la incertidumbre que rodea a la atribución, el uso de la mayor cantidad de líneas de evidencia posibles es una estrategia sensata, práctica y creíble.

Realizar una evaluación

En algunos casos, si las diversas líneas de evidencia apuntan hacia diferentes direcciones, habrá poco que pueda decirse con suficiente credibilidad sobre la contribución del programa. Si resulta crucial tener buena información sobre la atribución, entonces, la mejor estrategia es simplemente reconocer que uno no sabe y sugerir una evaluación con el fin de abordar la cuestión de la atribución. No obstante, en la mayoría de

los casos, si el programa realizó una contribución significativa, las diversas líneas de evidencia lo confirmarán.

Conclusión

En este trabajo, sostuvimos que lo que se necesita para tratar con la atribución usando la información de la medición de desempeño es explorar el asunto de manera sistemática y, al presentar los informes, dibujar una imagen creíble de la atribución para aumentar nuestro conocimiento de la contribución del programa. Debemos aceptar que, en la mayoría de los casos, lo que hacemos es medir con el objetivo de reducir la incertidumbre sobre la contribución realizada, no probar que esta se hizo.

Sugerimos emprender un *análisis de contribución* que examine y presente el mejor caso posible —una historia de desempeño creíble— para la atribución con la evidencia disponible. Básicamente, proponemos una *medición de desempeño basada en la teoría*, en la que, con el tiempo y el uso del análisis de contribución, se desarrolle y utilice una mejor comprensión de cómo funciona el programa para mejorar el desempeño futuro y para informar el desempeño pasado.

Una historia creíble de contribución

Mediante el uso del análisis de contribución, un caso razonable de que un programa marcó una diferencia supondría⁷:

- proporcionar una presentación bien articulada del contexto y de sus objetivos generales, junto con las estrategias para alcanzar estos fines;
- presentar una teoría del programa plausible, que conduzca a objetivos globales (no se refutó la lógica del programa, es decir, la evidencia es escasa o no contradictoria, y los supuestos subyacentes parecen seguir siendo válidos);
- describir las actividades y los productos del programa;
- destacar los resultados del análisis de contribución que indican una asociación entre las acciones del programa y los efectos observados;

- señalar que se descartaron las principales explicaciones alternativas para la ocurrencia de los efectos, como otros programas relacionados o factores externos, o que claramente tuvieron limitada influencia.

Si todo esto no es suficiente y hay demasiadas brechas en la historia, esto podría admitirse y aceptar la necesidad de una evaluación para una mejor comprensión de la contribución del programa. Es probable que el reconocimiento del problema y el conocimiento de otros factores en juego den paso a la recopilación de datos e información adicionales. El resultado será una mejor comprensión del programa y de cómo se espera que funcione, y quizá un rediseño para reflejar esta comprensión mejorada. Asimismo, una mejor información del desempeño proporcionará una demostración más creíble de los impactos del programa por medio de la medición de desempeño.

Notas

* Título original “Addressing attribution through contribution analysis: using performance measures sensibly”, *The Canadian Journal of Program Evaluation*, Ottawa, *Canadian Evaluation Society*, Vol. 16, 1, pp. 1-24. Publicación autorizada en inglés © The Canadian Journal of Program Evaluation. Traducción autorizada al español. Todos los derechos reservados.

Una versión anterior de este artículo se encuentra publicada en el sitio web de la Oficina del Auditor General de Canadá: <http://oag-bvg.gc.ca/domino/other.nsf/html/99_e.html>. Reproducido con la autorización del ministro de Obras Públicas y Servicios Gubernamentales, Ottawa, Ontario, Canadá K1 A 0S5, 2001.

¹ Este material está tomado de un ensayo conjunto de 1998 que fue confeccionado por la Oficina del Auditor General y por la Secretaría de la Junta del Tesoro.

² En general, los términos «efectos» y «resultados» se usan indistintamente. En sentido estricto, «resultados» incluye los productos (ver Figura 1) y, por lo tanto, es más amplio que «efectos». Sin embargo, gran parte de la literatura y algunos textos citados en este artículo usan el término «resultados» para significar «efectos» cuando la intención es clara. Si se hace referencia a los productos, se usará esta palabra.

³ En el caso de la presentación de informes, no estamos planteando que solo se debería recopilar o buscar evidencia que refuerza la afirmación del impacto del programa. Poder decir con confianza que no se sabe qué contribución hace el programa también es un conocimiento valioso. Por medio de las medidas de desempeño, tratamos

de reunir la mayor cantidad de evidencia práctica para comprender el alcance y la naturaleza de la contribución del programa y para respaldar esa afirmación.

4. Las ideas de esta sección fueron propuestas por Steve Montague de la Red de Administración de Desempeño de Ottawa.
5. El término y el ejemplo fueron desarrollados por Hugh McRoberts de la Oficina del Auditor General.
6. Aquí una advertencia es que si un experto individual tiene un interés particular en el programa, entonces sus opiniones deberán ser convenientemente descartadas.
7. Hendricks (1996) propone una lista similar.

Referencias

Auditor General of Canada. (1996). *Report to the House of Commons: Matters of special importance*. Ottawa: Author.

—. (1997). Report to the House of Commons. Capítulo 11. *Moving towards managing for results*. Ottawa: Author.

Banco Mundial. (1996). *Colombia: Paving the way for a results-oriented public sector*. Informe N.º 15300-CO. Washington, D. C.: Author.

Chen, H. T. (1990). *Theory-driven evaluation: A comprehensive perspective*. Thousand Oaks, California: Sage.

—. (1994). A panel on theory-driven evaluation and evaluation theories. *Evaluation Practice*, 15(1): 73-74.

Freeman, H. E. y Rossi, P. (1993). *Evaluation: A systemic approach*. Thousand Oaks, California: Sage.

Hatry, H. (1997). We need a new concept of accountability. *The Public Manager*, 26(3), 37-38.

Hendricks, M. (6 de noviembre de 1996). *Performance monitoring: How to measure effectively the results of our efforts*. Ensayo presentado en la Conferencia Anual de la Asociación Americana de Evaluación, Atlanta (Georgia).

Hudson, J., Mayne, J. y Thomlison, R. (Eds.). (1992). *Action-oriented evaluation: Canadian practices*. Toronto: Wall & Emerson.

Julian, D. A., Jones, A. y Devo, D. (1995). Open systems evaluation and the logic model: Program planning and evaluation tools. *Evaluation and Program Planning*, 18(4), 333-341.

Mayne, J. y Zapico-Goni, E. (Eds.). (1997). *Performance monitoring for public sector reform: Future directions from international experience*. New Brunswick, Nueva Jersey: Transaction Publishers.

Meekings, A. (1995). Unlocking the potential of performance measurement: A practical implementation guide. *Public Money & Management*, 15(4), 5-12.

Montague, S. (1998). *Build reach into your logic chart*. Disponible en <http://www.pmn.net/contributions/reachlog.htm> (Nota de la editora: sitio web no encontrado. Fecha de consulta: 28 de agosto de 2016).

Office of the Auditor General of Canada y Treasury Board Secretariat. (1998). *Modernizing accountability practices in the public sector*. Ottawa: Author.

Organización para la Cooperación y el Desarrollo Económicos. (1997). *In search of results: Performance management practices*. París: OECD.

President of the Treasury Board. (1997). *Accounting for results: 1997*. Informe Anual para el Parlamento. Ottawa: Treasury Board Secretariat.

United States General Accounting Office. (1996). *Observations on DOE's success stories report*. Testimonio ante la Subcomisión de Energía y Medio Ambiente, Comisión de Ciencias, Cámara de Diputados. GAO/T-RCED-96-133. Washington, D. C.: Author.

—. (1998). *Managing for results: Measuring program results that are under limited federal control*. GAO/GGD-99-16. Washington, D. C.: Author.

Weiss, C. (Invierno de 1997). Theory-based evaluation: Past, present and future. *New Directions for Evaluation*, 76, 41-55.

Wholey, J. S. (1983). *Evaluation and effective public management*. Boston, Massachusetts: Little, Brown.

Wholey, J., Hatry, H. y Newcomer, K. (1994). *Handbook of practical program evaluation*. San Francisco, California: Jossey-Bass.

La independencia de la evaluación en las organizaciones*

Roberto Picciotto

Introducción

«Queda mucho por hacer en el diseño de las instituciones para reconciliar los valores de autoridad y de responsabilidad» (Arrow, 1974: 77). La independencia de la evaluación es un tema aún desatendido en la literatura. Para ayudar a llenar el vacío, este artículo presenta un marco normativo sobre el papel de la evaluación independiente y de la autoevaluación en las organizaciones. Recurre a la teoría organizacional, así como a la propia experiencia del autor al frente del Grupo de Evaluación Independiente del Banco Mundial (1992-2002). ¿Qué es la evaluación independiente? ¿Cuál es su fundamento? ¿Puede diseñarse una función de evaluación interna e independiente para promover la rendición de cuentas organizacional —o está condenada a ser un ritual vacuo—? ¿Contribuye al aprendizaje organizacional —o produce un efecto desalentador que inhibe la adaptación a las cambiantes circunstancias—?

Antecedentes

Los gurús de las escuelas de negocios y los economistas institucionales todavía tienen que ponerse al día con el creciente papel de la evaluación en las organizaciones. En la comunidad de la evaluación, la dimensión de la gobernanza de la evaluación casi nunca se aborda, y la independencia de evaluación a menudo es vista como un estándar ético que los evaluadores deben cumplir en orden a generar análisis imparciales y hallazgos creíbles.

Esta estrecha e individualista concepción de la independencia de evaluación tiene consecuencias: impide el escrutinio de la independencia funcional y estructural e ignora las presiones manifiestas y encubiertas a las que los evaluadores están rutinariamente sujetos en el mundo real. Salvo raras excepciones, en las guías de evaluación no se reconoce la función de los comisionados como un determinante crítico de la calidad de evaluación (Picciotto, 2005). Por el contrario, se considera que la integridad de las evaluaciones depende ante todo de las competencias de los evaluadores, el sistema de aseguramiento de la calidad que se elija, la pertinencia de los métodos de evaluación y la transparencia de los procesos de evaluación.

La omisión de la independencia de la evaluación como un asunto de la gestión pública y corporativa es consistente con la visión generalizada de que las evaluaciones externas son independientes. Sin embargo, en la mayoría de los casos, la dependencia arancelaria y las restricciones contractuales constriñen seriamente la libertad de los evaluadores externos para ejercer un juicio imparcial. Asimismo, la falta de atención académica a la independencia de la evaluación puede deberse a la noción errónea de que la evaluación, al igual que la contabilidad y la auditoría, ha alcanzado el estatus de profesión hecha y derecha que goza de control autónomo sobre sus prácticas ocupacionales. Lamentablemente, esto no es así (Picciotto, 2011).

Por último, la evidencia de base sobre la relativa eficacia de la evaluación independiente frente a la autoevaluación es poco consistente, excepto en el ámbito del desarrollo internacional. Pocas organizaciones de los sectores público, privado y del voluntariado se han equipado con funciones de evaluación internas independientes. Aún hay un palpable desencanto público con las evaluaciones externas que no son en verdad independientes (Love, 1991), mientras que los acuerdos de evaluación interna son patentemente inapropiados por cuestiones de rendición de cuentas, dada la generalización del comportamiento egoísta en las organizaciones y su corolario: la renuencia para aceptar el fracaso (Vedung, 1997).

Es probable que en el futuro la evaluación independiente esté en el centro de atención en las organizaciones, dada la eterna preocupación por la reforma del sector público. La idealista retórica de resultados que indujo al

movimiento de la *nueva gestión pública* a prometer efectos utópicos a través de la delegación, la descentralización y la desregulación dio lugar a una postura realista que reconoce los límites del mercado basado en soluciones, la diversidad de contextos organizacionales y el valor de las mejoras graduales mediante mecanismos internos de ajuste (Pollitt y Boukaert, 2000).

¿Qué es la independencia de la evaluación?

Es generalizada la ambivalencia sobre la independencia de la evaluación entre quienes la practican. Solo los evaluadores que suscriben los principios de la evaluación democrática son acérrimos proponentes de la independencia de la evaluación. Y son una minoría. La mayoría de los evaluadores actúan como consultores. Le otorgan primacía a la utilización de los resultados de la evaluación y limitan su ambición a la provisión oportuna de conocimiento evaluativo a los responsables de tomar decisiones. También les preocupa el efecto desalentador y el aislamiento que la independencia puede producir.

En consecuencia, estos evaluadores son partidarios de un papel de asistencia estrictamente consultivo que no desafía los valores implícitos de sus empleadores ni los procesos que subyacen al desempeño organizacional¹. Sin duda, la independencia de la evaluación no es un fin en sí mismo. Pero, como el resto de este artículo busca demostrar, la lógica de la independencia de la evaluación se basa en la teoría organizacional (Arrow, 1974). Al fin y al cabo, las organizaciones existen para resolver las tensiones inherentes entre metas individuales y colectivas, para administrar flujos de información y para coordinar acciones a través de un nexo de contratos que mantienen los costos de transacción a raya. La evaluación independiente juega un papel clave en todas estas funciones.

La aplicación de los códigos de conducta al interior de las organizaciones implica autoridad, la cual es limitada, excepto en regímenes dictatoriales que no toleran ninguna oposición, y la libertad de acción gerencial no es absoluta. Esto es así porque la autoridad se tolera poco y no puede sostenerse en el tiempo a menos que sea legítima. A su vez, la legitimidad implica que las personas a quienes se les confió la dirección de la organización deben responder y rendir cuentas a una autoridad superior y a un público mayor.

Por lo tanto, responsabilizar a la autoridad requiere de mecanismos que ayuden a determinar si los errores en la toma de decisiones se debieron a circunstancias que escapaban al control de la organización o si los riesgos podrían haberse manejado diferente, de modo que los errores (con el beneficio de la retrospectiva) fueran innecesarios.

Mecanismos similares ayudan a informar la adaptación de los procesos internos y de los protocolos por medio del aprendizaje organizacional. Por supuesto, los errores deben evitarse en el propio proceso evaluativo, lo que significa que los procesos de evaluación deberían diseñarse y gestionarse, de manera que se evalúe el mérito y el valor de las actividades organizacionales de un modo justo, válido y preciso. Entonces, estos son los imperativos de una gobernanza sólida que la independencia de la evaluación ayuda a sostener.

La independencia de la evaluación contribuye a proteger la integridad del proceso de gestión. Mejora su credibilidad y proporciona nuevas perspectivas sobre las políticas y los programas que se evalúan (Mayne, 2008). Estas metas no pueden alcanzarse si las evaluaciones se establecen de manera que

- constriñen información, de manera que los productos de la evaluación no puedan tener ningún contenido crítico;
- controlan el contenido del programa de evaluación, de modo que no aporte nuevo conocimiento;
- demoran el proceso de evaluación (o la divulgación de resultados) hasta que las decisiones que podrían haber sido informadas por la evaluación hayan sido tomadas;
- inducen a los evaluadores a enfocarse en aspectos irrelevantes o marginales del programa o de la política que se evalúa.

La independencia y la calidad son sinérgicas

En palabras de Michael Scriven, la evaluación independiente «puede reducir ciertos tipos de sesgos (incluidos) (...) los conflictos de interés extremos, en los que el evaluador está "interesado" en el programa que se evalúa (...)

típico de muchos programas de monitoreo por agencias y fundaciones, en las que el monitor es en general el padrino del programa, a veces su inversor y casi siempre su defensor en la agencia» (Scriven, 1991: 192-193).

Habiendo dicho esto, la independencia en sí misma no garantiza la calidad de la evaluación: también se necesitan habilidades relevantes, métodos sólidos, recursos adecuados y transparencia. Las evaluaciones independientes pero incompetentes pueden ser engañosas, disruptivas y costosas. Por su parte, la calidad de evaluación sin independencia pierde credibilidad. Esta es la razón por la cual las personas y las entidades con motivos para temer el resultado de una evaluación frecuentemente pondrán en duda su independencia.

En entornos de trabajo abiertos y responsables, la independencia de la evaluación produce confianza, protege el proceso de aprendizaje e induce a administradores y partes interesadas del programa a centrarse en los resultados. Junto con la garantía de la calidad y los lineamientos éticos, la independencia propiamente definida debería ser una característica reconocida de credibilidad y de fiabilidad para los sistemas de evaluación internos y esto debería reflejarse en los lineamientos de evaluación que confeccionan las asociaciones de evaluación.

La independencia es una competencia

Ante todo, la independencia de la evaluación es una característica de la gobernanza corporativa, aunque tiene implicancias para las competencias de los evaluadores. En la evaluación como en la ciencia, el conocimiento y la práctica no bastan para alcanzar la excelencia. La disposición y las actitudes también son relevantes. Los evaluadores se caracterizan por su curiosidad, escepticismo y sed de evidencia. Asimismo, la evaluación requiere valentía e imparcialidad, reflejadas en un enfoque equilibrado y seguro de sí mismo que no vacila en señalar problemas ni déficits de desempeño, pero que además reconoce éxito y logros. De este modo, el Libro Amarillo de la Oficina de Rendición de Cuentas del Gobierno de los Estados Unidos, que establece los criterios para auditorías y evaluaciones, identifica «una actitud y una apariencia independientes» como características deseables.

Asimismo, de acuerdo con el Manual de la Federación Internacional de Contadores (2003), la independencia de criterio «permite aportar una opinión sin que esta se vea afectada por las influencias que comprometen el juicio profesional, lo que hace que una persona actúe con integridad y ejerza la objetividad y el escepticismo profesional», mientras que la independencia en apariencia es «evitar hechos y circunstancias que son tan importantes que un tercero sensato e informado, con conocimientos de toda la información relevante, incluidas las salvaguardias, concluiría razonablemente que la integridad, la objetividad o el escepticismo profesional se han visto comprometidos». Ambos significados son reconocidos como capacidades de los evaluadores por la Sociedad Europea de Evaluación (2011).

Los evaluadores independientes necesitan protección

El carácter independiente importa, pero nadie puede razonablemente esperar que los evaluadores actúen de forma independiente en contextos organizacionales que ponen sus carreras, su reputación y su sustento en riesgo. Necesitan protección adecuada para llevar adelante su demandante y estresante mandato. Se requieren salvaguardias institucionales especiales. En concreto, los evaluadores deben estar protegidos contra las amenazas externas a su imparcialidad. Se les debería otorgar total acceso a la información que precisan para realizar su trabajo. La inmunidad de la captación de cualquiera de las partes que comparte la responsabilidad de la administración del programa es fundamental.

¿Cómo se otorga y se protege la independencia de la evaluación?

Si a los evaluadores externos se les encarga una evaluación, un contrato regula el proceso. Los lineamientos facultativos emitidos por las sociedades de evaluación también orientan el trabajo. Sin embargo, el grado real de independencia, en gran medida, depende de si la comisión de evaluación está libre de lealtades y desprovista de intereses sobre la política o el programa a evaluar, y si intenta genuinamente descubrir si una política o programa funcionan y de no ser así, averiguar el porqué y cómo podría mejorarse.

Así la independencia de la evaluación es una cuestión impredecible en circunstancias en las que la organización que contrata una evaluación no tiene salvaguardias precisas especificadas que garanticen la independencia y la transparencia para el proceso de evaluación. ¿Qué son estas salvaguardias? Fueron codificadas para casos en los que la evaluación es realizada o es contratada por una unidad que está integrada en una organización. Específicamente, el Grupo de Cooperación de Evaluación (2010) identificó cuatro dimensiones fundamentales de la independencia de la evaluación sobre la base de una cuidadosa revisión bibliográfica y de décadas de experiencia en evaluación independiente, en el ámbito del desarrollo internacional: (a) independencia organizacional; (b) independencia de comportamiento; (c) elusión de conflictos de interés; (d) protección contra influencia externa. Se proporciona más información en el Anexo.

La independencia organizacional asegura que el personal de la evaluación no sea controlado ni influenciado por los que toman las decisiones, quienes son responsables por las actividades que se evalúan y que, dentro de restricciones éticas y legales, tienen acceso completo a la información que necesitan para cumplir su mandato. La independencia de comportamiento mide hasta qué punto la unidad de evaluación puede establecer —y está dispuesta a hacerlo— su programa de trabajo, elaborar informes de alta calidad e inflexibles y divulgar sus hallazgos ante la junta sin restricciones impuestas por la gestión.

Las salvaguardias en materia de conflictos de interés garantizan que las relaciones y las consideraciones personales y profesionales actuales, del futuro inmediato o anteriores no puedan influenciar los juicios de los evaluadores ni crear la apariencia de falta de objetividad. La protección contra la influencia exterior mantiene libre la función de la evaluación para fijar sus prioridades, diseñar sus procesos y productos, alcanzar sus juicios y administrar sus recursos humanos y presupuestarios sin interferencia de la administración.

Estos criterios de independencia de la evaluación están interrelacionados. Todos ellos son importantes. La protección contra la influencia exterior es el objetivo final de la independencia organizacional. Los conflictos de

interés son frecuentes ante la ausencia de independencia organizacional. La independencia de comportamiento es una función de la independencia organizacional, así como la elusión de conflictos de interés y la protección contra influencias externas².

La externalidad no garantiza la independencia

A esta altura, debería quedar claro que la evaluación externa no puede ser equiparada con la independencia de la evaluación. La dependencia arancelaria es una amenaza muy importante a la integridad del proceso de evaluación. El juicio de los evaluadores externos se ve inhibido o amenazado si sus servicios son retenidos por los administradores a cargo de las actividades que se evalúan. En cambio, las unidades internas de evaluación que son financiadas y controladas por la máxima autoridad de gobernanza están mejor protegidas contra la interferencia de la dirección, debido al sistema de control y equilibrio asociado al papel estatutario de supervisión que las juntas directivas deben desempeñar.

Al mismo tiempo, las unidades de evaluación interna independientes respaldadas por un mandato adecuado y con evaluadores familiarizados con el terreno organizacional gozan de una relativa proximidad con los programas que se evalúan. De este modo, las evaluaciones realizadas por unidades internas independientes que reportan a las juntas directivas tienen muchas más probabilidades de superar las «asimetrías de la información», mientras protegen la objetividad del proceso evaluativo.

La independencia no es aislamiento

La prueba de fuego de la independencia y la calidad está vinculada con la medida en la que las evaluaciones están dotadas de

- *criticalidad*: capacidad y predisposición para juzgar el desempeño de manera objetiva y transparente;
- *adicionalidad*: contribución distintiva a la creación o divulgación del conocimiento operativo;

- *oportunidad*: entrega de hallazgos de las operaciones de evaluación y lecciones con suficiente antelación como para informar la toma de decisiones;
- *materialidad*: foco deliberado en temas y cuestiones de sustancial relevancia para la eficacia del desarrollo.

Algunos evaluadores sostienen convincentemente que no tener conexión ni experiencia compartida con los usuarios previstos de los hallazgos de la evaluación restringe el acceso a la información, demuestra resistencia e inhibe el aprendizaje. No cabe duda de que las evaluaciones precisas e imparciales combinan la objetividad intelectual con la empatía y la comprensión. La capacidad para involucrarse con las diversas partes interesadas y para asegurar su confianza, mientras se conserva la integridad del proceso de evaluación es un sello de profesionalismo. La impugnación de los hallazgos de la evaluación independiente y los desacuerdos respetuosos y ejemplares sobre las recomendaciones contribuyen a una saludable cultura de aprendizaje organizacional.

Por el contrario, las actitudes conflictivas y los enfoques de «señalar y avergonzar» rompen el contacto con los responsables de tomar las decisiones, restringen el acceso al conocimiento tácito, inhiben los intercambios profesionales y aumentan la resistencia a la adopción de las recomendaciones de la evaluación. Estos conducen al aislamiento, a la falta de apalancamiento intelectual y a un efecto desalentador en materia de aprendizaje organizacional. Por esto, se establecen rendimientos decrecientes cuando la independencia de evaluación adopta formas extremas y antagonistas.

La evaluación independiente puede ayudar a conectar el conocimiento del programa con la estrategia

La teoría organizacional sostiene que la creación de valor público por parte de una organización requiere capacidades operativas alineadas con las demandas de su *entorno de autoridad* (Moore, 1995). Por lo tanto, satisfacer las necesidades de rendición de cuentas y de conocimiento de las partes interesadas es central para el diseño de una función receptiva de la

evaluación. Esto es especialmente importante en el sector público y en el voluntariado, puesto que el principal mecanismo de restablecimiento para el mal desempeño en la esfera pública es la opción de «voz», mientras que en el sector privado la opción de «salida», ejercida a través de los mecanismos de mercado competitivo, es el modo en que las partes interesadas expresan su disconformidad (Hirschman, 1970).

Los accionistas descontentos pueden vender sus acciones y los consumidores insatisfechos pueden cambiar de una marca a otra o de un proveedor a otro. Otra diferencia entre el mercado económico y el político es que los proveedores privados atienden necesidades bastante homogéneas que se traducen fácilmente en mediciones de mérito y valor. En concreto, las empresas privadas acceden a la información sobre su desempeño (pruebas de mercado y ratios financieros), mientras que en la esfera pública, los servicios sociales atienden variados intereses que deben satisfacerse de un ciclo electoral al otro.

En el intervalo entre los ciclos electorales, los mecanismos de retroalimentación deben considerar las opiniones de las distintas partes interesadas y lograr compensaciones adecuadas entre ellas. La acción pública impacta y trabaja por medio del sector privado y de la sociedad civil. Esto significa que los responsables de formular las políticas y los administradores de los programas sociales no pueden crear valor mediante la dependencia exclusiva de factores que pueden controlar total o directamente. En particular, la opinión pública es el elemento vital del gobierno. Para las organizaciones del sector público, esto pone especial énfasis en las relaciones comunitarias y en las campañas de información que pueden desencadenar el voluntariado y el respaldo público. Por consiguiente, para la evaluación, el entorno operativo externo debería ayudar a generar indicadores que informan las decisiones estratégicas. Por lo tanto, una función de evaluación independiente está bien ubicada para validar indicadores y procesos que conectan a las organizaciones del sectores público y del voluntariado con los ciudadanos que sirven. La retroalimentación de los diversos distritos electorales está garantizada por una diversidad de instrumentos de evaluación de la participación ya consagrados (grupos focales, encuestas, etcétera) y, cada vez más, por medio de las tecnologías de las redes sociales.

En resumen, la evaluación no puede alcanzar su potencial a menos que conecte eficazmente a sus gestores, a las máximas autoridades que gobiernan la organización y la sociedad en su conjunto. Pero al hacerlo, debería conservar su objetividad, ejercer plena libertad de investigación y resistir la captación. Esto es un imperativo, ya que sin independencia, el público percibe que la evaluación está supeditada a intereses creados: para que el conocimiento de la evaluación sea creíble, legítimo y valioso, la evaluación debe ser funcional y estructuralmente independiente.

La evaluación independiente mejora la rendición de cuentas organizacional

Tanto en el sector público como privado, los administradores tienen el mandato de asegurar un alto valor para el paquete de activos asignado a su cuidado (Moore, 1995). Deben demostrar a quienes controlan su financiamiento que están administrando los recursos que les confiaron responsable y eficazmente. Son responsables de alcanzar resultados. Sin embargo, las organizaciones públicas y privadas se aseguran los fondos necesarios para realizar sus actividades de distintas maneras. En el sector privado, los ingresos provienen de la venta de bienes y servicios a los consumidores individuales —o de los mercados de capitales privados que determinan el valor prospectivo de dichas ventas—.

En el sector público, la máxima fuente de fondos es el contribuyente. Compete a los funcionarios públicos y, en última instancia, a los políticos efectuar elecciones colectivas sobre cómo asignar los recursos públicos. En consecuencia, la clave de acceso a los recursos en el sector público es una narrativa válida y autorizada sobre la creación de valor público juiciosamente dirigida a las personas con autoridad. La dificultad de medir el valor social explica por qué esas medidas simples y económicas de producción y los coeficientes presupuestarios, más que los resultados y los impactos han dominado tradicionalmente la administración del sector público.

Dichos indicadores poseen amplias y bien conocidas desventajas. No miden resultados y pueden ser manipulados con facilidad. Por lo tanto, la información que los administradores del sector público brindan sobre su trabajo requiere una validación independiente: la evaluación

independiente del sector público es el equivalente a la auditoría de cuentas del sector privado. Por este motivo es evidente la necesidad de una vinculación estrecha entre evaluación independiente y rendición de cuentas organizacional. Si la evaluación no logra mejorarla, será por falta de validez debido a su mala calidad y/o porque no puede satisfacer los criterios que la hacen verdaderamente independiente.

¿Qué es el aprendizaje organizacional?

En una organización de aprendizaje, los empleados se involucran en un aprendizaje permanente. El aprendizaje solo puede establecerse con confianza, mediante cambios en el comportamiento. Pero no todos esos cambios derivan de la adquisición de conocimiento o de la experiencia. El condicionamiento a través del temor y la recompensa, del respecto por la autoridad y del deseo de amoldarse, entre otros, también afecta las acciones individuales. El aprendizaje solo se produce si nuestros pensamientos y acciones fueron moldeados a partir de una clara comprensión del entorno en el que vivimos.

El aprendizaje no es automático. La intervención externa es crítica para la superación personal y la comprensión profunda. El aprendizaje desafía ideas y preconceptos vigentes. Puede ser provocado por eventos inesperados o por amenazas externas. También puede ser nutrido por instrucción, entrenamiento, o educación formal. Da lugar a nuevos discernimientos e incentiva la adquisición de nuevos conceptos y modelos mentales que sirven de útiles guías de acción. Esto tiene implicancias significativas para la eficacia organizacional.

Sin el conocimiento, experiencia y habilidades apropiados, las personas libradas a su suerte presentan persistentes déficits de comprensión basados en ideas erróneas, generalizaciones exageradas, interpretaciones equívocas de la evidencia e incapacidad para formular puntos de vista convincentes, entre otros (Perkins, 1993). La ciencia cognitiva, la psicología educacional y la experiencia práctica demuestran que el aprendizaje no es automático y requiere interacción con agentes experimentados, conocedores y externos. Esta es la razón por la cual la eficacia organizacional depende, en gran medida, de

sólidas políticas de recursos humanos, una dimensión crítica de la eficacia organizacional (...), pero la historia no termina aquí.

El aprendizaje organizacional tiene su propia dinámica

Las políticas públicas y los programas son diseñados y están implementados por las organizaciones. Los responsables internos de tomar las decisiones no están libres de restricciones, sino condicionados a cumplir con «reglas de juego». Estas reglas están moldeadas por restricciones legales, procedimentales y de tradición, así como por costumbres ya consagradas que favorecen los modos habituales de hacer las cosas. Las normas presupuestarias, las prácticas en materia de recursos humanos y los procedimientos operativos generan poderosos incentivos.

Por consiguiente, el aprendizaje organizacional no es igual al aprendizaje individual. Sin duda, no pueden separarse, ya que las personas son las que hacen que una organización funcione. Pero las decisiones organizacionales y estratégicas son decisiones colectivas moldeadas por jerarquía, protocolos y antecedentes. A través de la evaluación independiente, el aprendizaje organizacional y el aprendizaje individual se conectan. Por este motivo, la proposición generalizada de que la independencia en la evaluación no conduce al aprendizaje es errónea. Las organizaciones bien gobernadas tratan al aprendizaje organizacional como una responsabilidad de la dirección.

Dilemas de la acción colectiva

En las organizaciones, particularmente las grandes, el comportamiento vinculado con el *free riding* es una elección racional para los miembros de un grupo individual (Olson, 1971). Este problema solo puede mitigarse con incentivos compensatorios que fomenten el esfuerzo disciplinado hacia las metas corporativas y que sancionen el comportamiento oportunista. Esto explica por qué se necesitan sistemas de control en organizaciones bien manejadas. Sin ellos, las políticas y los estándares pueden ser ignorados.

Asimismo, las organizaciones deben superar los obstáculos inherentes al acceso a la información asimétrico. El escarpado volumen de información que debe procesarse para una toma de decisiones eficaz y oportuna

requiere delegación de autoridad. A su vez, esto incrementa los problemas de principal-agente y de coordinación que no pueden resolverse sin «reglas de juego» diseñadas para minimizar los costos de transacción internos. Los sistemas de mando y control son necesarios, aunque cuando son estrictos e inflexibles, socavan la creatividad, la innovación o la capacidad de respuesta a las muy diferenciadas necesidades de las partes interesadas. Por consiguiente, los cambios periódicos en el entorno de control son necesarios para ayudar a la organización a adaptarse a un entorno en evolución. Una vez más, aquí entra en juego la evaluación independiente.

Ingreso de la evaluación

La dependencia de la trayectoria comúnmente asociada con estándares fijos y normas burocráticas socava la capacidad de respuesta organizacional. Por lo tanto, la garantía de calidad es necesaria para el aprendizaje organizacional, pero debe realizarse en «tiempo real», debería diseñarse para superar las asimetrías de la información a un costo razonable y debería ser susceptible a cambios en el entorno operativo y de autoridad. Por consiguiente, los protocolos operativos y los estándares de garantía de calidad deberían revisarse continuamente mediante procesos evaluativos.

En resumen, los sistemas de control corporativo son esenciales para asegurar la alineación de las acciones del personal con las metas de la empresa, especialmente en las organizaciones grandes. Pero también pueden contribuir a que la toma de decisiones sea lenta, costosa e inflexible. En particular, los sistemas de control de calidad fuertes que se ejercen desde lo más alto de la jerarquía pueden tener un efecto no intencionado de generar miedo al castigo y su desagradable corolario: el ocultamiento de errores. De este modo, los protocolos de negocio detallados pueden tener una consecuencia no deseada: aprendizaje empresarial e innovación limitados. Básicamente, este es el motivo por el cual la evaluación independiente es necesaria para «sacudir las cosas», especialmente en las organizaciones grandes y exitosas: la función de una evaluación independiente proporciona un contrapeso muy necesario dentro de la estructura de los incentivos empresariales al cambiar el foco de la gestión de los insumos a los resultados.

Los límites del liderazgo carismático

Los líderes inteligentes, seguros de sí mismos, carismáticos y sensibles a las demandas en evolución del entorno operativo pueden ayudar a minimizar los riesgos de impacto negativo asociados con los rígidos sistemas de control corporativo. A pesar de su enérgicamente autoproclamado compromiso para con la rendición de cuentas, la transparencia y el aprendizaje, esos líderes fuertes tienden a ser reacios a reconocer errores. A menudo, equiparan la lealtad hacia ellos con la lealtad a la organización y son propensos al uso de los medios y las técnicas de las comunicaciones modernas para mitigar los vigorosos efectos de una transparencia genuina. Las sólidas estructuras de la gobernanza, la hábil supervisión de la junta y el escrutinio de la sociedad civil pueden ayudar a traspasar la niebla de las relaciones públicas que los líderes carismáticos son propensos a crear, pero sin la evaluación independiente resulta difícil para la máxima autoridad de la gobernanza de una organización pública grande y compleja —y mucho más para la ciudadanía— descubrir qué sucede y generar una narrativa conflictiva.

Evaluación para el aprendizaje organizacional

La mayoría de las organizaciones aún invierten poco en evaluación. Sus líderes no reconocen el papel positivo que la evaluación puede jugar en la mejora del desempeño organizacional. La realidad es que una cultura de cumplimiento no es propicia para una sólida gestión del riesgo en el mundo real. Los mecanismos tradicionales de supervisión tienen su espacio, pero, en general, inhiben la adaptación ágil al cambio. De esto se trata el aprendizaje organizacional guiado por la evaluación independiente. Facilita el cambio influenciando cómo la organización clasifica, procesa, almacena y usa la información necesaria para la toma de decisiones.

Transformar información en conocimiento requiere el ejercicio del juicio profesional. Si el mandato organizacional es complejo, la especialización del personal es un imperativo, aunque los sesgos disciplinarios pueden distorsionar los sistemas de garantía de calidad y limitar las innovaciones. Se necesitan múltiples perspectivas y comunicaciones internas efectivas. Si bien la especialización favorece la pericia y la eficiencia, la mentalidad de silo puede inhibir el pensamiento lateral que puede ser la clave del éxito

cuando la flexibilidad y la creatividad son muy solicitadas. La evaluación está bien equipada para enfrentar estos riesgos.

Los roles generalistas promueven flexibilidad e interdependencia, pero una excesiva tolerancia a la ambigüedad puede conducir a una descuidada toma de decisiones. Entonces, se deduce que las organizaciones de aprendizaje son las que dominaron los dilemas de la acción colectiva vinculados con el manejo de la información y la toma de decisiones en los grupos. Especialmente en entornos operativos volátiles, el debate vivaz dentro de la organización, la sensibilidad a necesidades y restricciones de las partes interesadas, la apertura a nuevas ideas, la buena disposición a soltar estrategias anticuadas, la ejecución ágil, las soluciones innovadoras y el gestión del riesgo (en lugar de evitarlo) son, en general, más importantes que el cumplimiento de políticas y normas establecidas.

Un cambio organizacional es, en consecuencia, un proceso desafiante dados los obstáculos inherentes en las tensiones entre la continuidad de la política y la adaptación a circunstancias cambiantes, la centralización del establecimiento de metas y la descentralización de la toma de decisiones, la especialización y la apertura a diversas disciplinas. Por lo tanto, estas son las paradojas que subyacen a la necesidad de contar con mecanismos explícitos diseñados para nutrir la innovación y el aprendizaje organizacional. La evaluación independiente es uno de esos mecanismos.

El aprendizaje a partir de la experiencia

Integrada a los procesos de los negocios y concentrada en los resultados, la evaluación independiente puede ayudar a resolver los dilemas de la acción colectiva que acosan a las grandes organizaciones jerárquicas. ¿De qué forma la evaluación induce el cambio organizacional positivo? Una vasta literatura se ha acumulado sobre este tema. Como era de esperar, el consenso general es que la calidad de la evaluación es un imperativo. Una evaluación engañosa e injusta puede provocar un enorme daño organizacional. El juramento hipocrático («Primero, no hacer daño») se aplica a los evaluadores al igual que a los médicos. La mala calidad de una evaluación perjudica la disciplina de la evaluación.

Habiendo dicho esto, la influencia del informe de una evaluación individual solo se correlaciona ligera y parcialmente con su rigor y calidad. La destilación experta e independiente de experiencia no garantiza el uso de las lecciones extraídas. Otros factores (relevancia, oportunidad, métodos de divulgación, etcétera) importan más. Por lo general, la asimilación de las lecciones de la evaluación está sujeta a la compleja dinámica política y administrativa (Weiss, 1998). En el uso de la evaluación, es más relevante el contexto organizacional y cultural en el que esta se organiza y si es auspiciosa para el aprendizaje organizacional. En otros lugares, el proceso de evaluación es apto para generar reacciones complejas, sutiles y con frecuencia demoradas.

Por este motivo, la evaluación independiente no debería juzgarse únicamente por sus resultados instrumentales. Para estar seguros, los evaluadores independientes deberían diseñar, planificar y difundir sus productos, de manera que faciliten el uso de la evaluación, aunque para garantizar la objetividad, deberían operar fuera del alcance de la toma de decisiones. Independencia significa que no los pueden responsabilizar por la utilización de los hallazgos de la evaluación. Esta responsabilidad recae directamente en los responsables de formular las políticas y en los administradores del programa.

La evaluación tiene buenas características públicas

Cuando se resisten a la evaluación independiente, los mecanismos compensatorios de control, que buscan vincular linealmente las acciones de la administración con las recomendaciones específicas de la evaluación, tienden a ser ineficaces, al menos a corto plazo. No obstante, los hallazgos que se dejan a un lado en primera instancia pueden promover el debate interno, intensificar las presiones de las partes interesadas e inducir el escrutinio público que, a medio y largo plazo, inevitablemente provocarán un cambio positivo. La aceptación paulatina, tácita y con recelo de las prescripciones de la evaluación no es infrecuente.

Puede dar como resultado reformas parciales y a veces ocultas, como por ejemplo, cuando los agentes de cambio dentro de la organización deciden actuar en su propia esfera de influencia. Así, las ideas generadas por una

evaluación pueden llegar a buen puerto después de una crisis o cuando una nueva administración toma las riendas de la organización. En consecuencia, con el tiempo, se pueden inducir cambios de políticas de orden superior que involucren diversos y poderosos intereses. Bastan solamente algunos de esos casos para justificar las inversiones empresariales en evaluación.

Además, el uso *instrumental* a corto y medio plazo es solo uno de los potenciales beneficios que ofrece la evaluación. El uso *conceptual* puede ser más significativo que el uso instrumental en una organización resistente al aprendizaje. Con el tiempo, un proceso sólido de evaluación puede influir en el personal del programa para afilar el diseño de la política y del programa, y puede promover ideas que probablemente mejoren la implementación. La evaluación también puede *empoderar a los agentes de cambio interno* confirmando sus percepciones o sacando a la luz realidades incómodas que antes se escondieron debajo de la alfombra. Por último, las evaluaciones pueden tener efectos positivos fuera de la organización mediante la *creación de conocimiento* y las contribuciones al entendimiento público.

Habiendo dicho esto, un fundamento válido para invertir en evaluación se encuentra en la noción práctica de que los cambios organizacionales logrados oportunamente en un entorno turbulento y demandante no deberían dejarse librados a la serendipidad y la casualidad. El uso instrumental de la evaluación, es decir, la aplicación directa de las válidas lecciones aprendidas de los programas anteriores en el diseño e implementación de nuevos programas, presenta evidentes ventajas. Estos beneficios se aprovechan al máximo solo si la evaluación se integra a la cultura organizacional.

Del aprendizaje unidireccional al bidireccional

El pasado no siempre es un prólogo. Los logros organizacionales anteriores no necesariamente presagian un éxito continuo. De hecho, el éxito organizacional puede ser una maldición, ya que, por lo general, conduce a la complacencia y a la resistencia al cambio porque «si no está roto, no debería arreglarse». Con el fin de dominar la adaptación exitosa a circunstancias en evolución, las organizaciones de aprendizaje

(a) aseguraron las competencias correctas y nutrieron las habilidades adecuadas a través del aprendizaje individual relevante; (b) resolvieron constructivamente las tensiones que inevitablemente surgen entre continuidad y cambio organizacional, flexibilidad y control, cohesión interna y capacidad de respuesta para con las partes interesadas.

Tanto para las personas como para las organizaciones, detectar y corregir desviaciones de las normas constituye un *aprendizaje unidireccional*. Esta dimensión de la eficacia organizacional se centra en «hacer las cosas bien» y es el terreno privilegiado del monitoreo. Asimismo, conducir las evaluaciones de metas, políticas y protocolos que subyacen al comportamiento organizacional actual para «hacer las cosas bien» es un *aprendizaje bidireccional* que ocurre cuando se detecta y se corrige un error en formas que involucran la modificación de normas, políticas u objetivos subyacentes de una organización (Argyris, 1977). De esta función vital se trata la evaluación en una organización.

Por un lado, el aprendizaje unidireccional es de crítica importancia, puesto que ninguna organización grande puede funcionar eficientemente sin protocolos ni reglas que minimicen los costos de transacción internos, permitan la delegación, limiten los requerimientos de coordinación y aseguren la calidad del servicio, la puntualidad de la entrega y la capacidad de respuesta para con los clientes. Moral fuerte, trabajo en equipo cohesivo y lealtad institucional son los ingredientes críticos de la excelencia organizacional.

Por el otro, dada la complejidad y la volatilidad asociadas a un sistema global cada vez más interconectado y turbulento, las estrategias empresariales y los principios y estándares operativos tienen una corta vida útil. Necesitan ser adaptados periódicamente por medio del aprendizaje bidireccional para reflejar los cambios de los entornos operativos y de autoridad. De este modo, se debe alcanzar un prudente equilibrio entre los controles (aprendizaje unidireccional) y la evaluación (aprendizaje bidireccional).

La evaluación independiente como aprendizaje bidireccional

El aprendizaje unidireccional implementado a través de los mecanismos de auditoría y de supervisión es suficiente cuando los objetivos organizacionales están abiertamente respaldados, bien definidos, tienen plazos concretos, son medibles e inequívocos, y el entorno operativo es estable. En esas circunstancias, los modelos de desempeño que se centran en medidas fácilmente verificables proporcionan una conexión clara entre acciones organizacionales y resultados previstos. En este contexto, los modelos tradicionales de administración que enfatizan el control de los procesos que vinculan insumos organizacionales con productos, resultados e impactos están totalmente justificados, y el monitoreo disfruta de un lugar de honor en la caja de herramientas de gestión.

Pero a menudo las organizaciones grandes y complejas deben satisfacer múltiples objetivos en materia de políticas, por lo que, en general, operan en contextos que se caracterizan por su complejidad e inestabilidad. En consecuencia, el aprendizaje bidireccional también es necesario, y los modelos lineales que dependen de lograr consistencia entre metas predeterminadas y rígidos instrumentos de políticas obstaculizan la eficacia organizacional. Es preferible un enfoque de sistemas orientado a una administración ágil y pragmática.

Los enfoques de administración organizacional (y modelos de evaluación) que se concentran sobre todo en el desempeño interno y en la evaluación son apropiados para situaciones que requieren acciones conjuntas de diversos asociados, acuerdos entre ellos, resoluciones creativas de las tensiones entre prioridades rivales en materia de políticas y salvaguardias adecuadas para los segmentos más pobres y más vulnerables de la sociedad.

En esas circunstancias, la evaluación debería abarcar a todos los asociados y las evaluaciones basadas en metas periódicas deberían complementarse con evaluaciones de impacto libres de metas mediante el uso de perspectivas analíticas de valoración que sean realistas, éticas y legítimas. En particular, para que las evaluaciones sean éticas deberían amplificar las voces de la parte más débil de la relación. Deberían confiar

en indicadores de éxito verificables que estén acordados con todas las partes interesadas.

También es necesaria una perspectiva global para una organización internacional. Las intervenciones principalmente enfocadas en las intervenciones de un solo país pueden tener consecuencias no previstas si no logran tener en cuenta sus potenciales consecuencias para el sistema internacional.

Alcanzar un buen equilibrio

Sin duda, la dependencia exclusiva del aprendizaje bidireccional puede conducir a altos costos de transacción internos y a un caos organizacional. Puede inducir la agitación constante en la organización, agendas de políticas que se modifican con rapidez y prioridades cambiantes que terminan por socavar la eficacia organizacional. Por su parte, el aprendizaje unidireccional en sí mismo no ayuda a resolver las paradojas inherentes en la administración de las organizaciones grandes y complejas que operan en un contexto cambiante.

Esto sucede porque el cumplimiento de las normas es para lo que los empleados están condicionados a trabajar, aun después de que las cuestiones para las que esas normas fueron diseñadas ya no cumplen su propósito. En consecuencia, la importancia relativa del aprendizaje bidireccional (y, por lo tanto, de la evaluación) aumenta en entornos operativos inestables sujetos a rápidos cambios. Darle más énfasis al aprendizaje bidireccional (y a la evaluación) en las organizaciones tradicionalmente comprometidas con un aprendizaje unidireccional es especialmente difícil de inducir si han sido exitosas o si las normas actuales reflejan las demandas de un entorno de autoridad que está desfasado de la realidad.

En la mayoría de los casos, los intereses que empoderaron a las unidades corporativas para diseñar normas organizacionales harán su máximo esfuerzo por defenderlas y dotarán la jerarquía organizacional para hacerlas cumplir y sancionar a quienes crean en otra cosa. En las organizaciones orientadas al cumplimiento, los empleados que plantean preguntas fundamentales sobre las políticas, los protocolos y los estándares

existentes no ascienden en la pirámide de la organización. Son rechazados por sus colegas y su comportamiento puede percibirse como desleal. Esto explica por qué el hecho de denunciar irregularidades es tan raro en las organizaciones y también por qué se necesitan un guardián interno y unidades de evaluación independiente, los que deben proteger a sus miembros del castigo para ser eficaces.

Nexo entre evaluación independiente y autoevaluación

A esta altura debería quedar claro por qué el aprendizaje unidireccional a través del monitoreo es un proceso continuo respaldado por las unidades de control interno, mientras que las intervenciones de aprendizaje bidireccional manejadas por las unidades de evaluación independiente no necesitan ser tan frecuentes. Desde la perspectiva del aprendizaje organizacional, el desafío central de la evaluación independiente es ayudar a influir en la adaptación organizacional sin interferir con la administración. La evaluación independiente debe encontrar formas de influenciar otras funciones corporativas y operativas, al mismo tiempo que funciona fuera el alcance de ellas. Precisamente porque la evaluación independiente es selectiva e intermitente, necesita el apalancamiento de la autoevaluación para ser totalmente efectiva. Por supuesto, esto implica que la relación entre las dos funciones se basa en el respeto mutuo y en el profesionalismo.

La evaluación independiente necesita de la autoevaluación

Por un lado, el aprendizaje bidireccional debería estar garantizado por unidades de evaluación totalmente protegidas de los intereses creados. Estas pueden tener un impacto útil tan solo a través de la generación de informes válidos, confiables e intransigentes. Por otro lado, dichas unidades no deberían ser retiradas de la organización por ahora, de modo que un clima de desconfianza y de desacuerdo las desconecte totalmente del ciclo operativo y de políticas. Para alcanzar su máximo potencial, la evaluación independiente necesita apalancar su impacto con la autoevaluación. Ese es el puente que conecta los hallazgos de la evaluación con las decisiones de la administración.

La autoevaluación goza de ventajas únicas

En igualdad de condiciones, la autoevaluación es más eficaz que la evaluación independiente, debido a que sus costos de generación de información y de transferencia son más bajos. Es definitivamente más oportuna, por ejemplo, cuando se imponen periodos de enfriamiento a los temas seleccionados para la revisión independiente. La autoevaluación genera considerables beneficios mejorando el diseño de programas y políticas por medio del aseguramiento de la calidad, de la mejora del monitoreo y del seguimiento de los resultados.

Por último, es más probable que los hallazgos de la autoevaluación pertenezcan a los responsables de tomar las decisiones, dado que son autogenerados. Por lo tanto, la autoevaluación tiene tantas ventajas con respecto a la evaluación independiente que debería adoptarse y hacer uso de ella en la máxima medida posible. Sin embargo, padece sus propias limitaciones y, por muchas razones, necesita de la evaluación independiente para funcionar mejor.

La autoevaluación necesita de la evaluación independiente

Los administradores de los programas y los responsables de formular las políticas, a menudo, carecen de competencias de evaluación. Pueden tener intereses y preocupaciones diferentes de los de los beneficiarios del programa. Por lo general, sucumben ante los saltos de fe que conducen a decisiones imperfectas o excesivamente riesgosas. Se pueden ver tentados a seleccionar evidencia que respalde sus preconceptos. La forma de pensar de los evaluadores independientes insta a los autoevaluadores a ser más escépticos y reflexivos sobre los supuestos subyacentes al diseño de políticas y programas que confeccionan e implementan.

Los evaluadores independientes inducen a los autoevaluadores a pensar más sobre qué intenta lograr la organización, a consultar más sistemáticamente a las partes interesadas, a buscar un consenso resiliente sobre las metas del programa y a movilizar los recursos dispersos y las energías hacia las metas del programa. Además, la evaluación independiente resguarda la rendición de cuentas cuando la autoevaluación es débil. La supervisión

profesional de la autoevaluación por parte de la evaluación independiente mejorar la eficacia de la primera de formas similares a aquellas que hacen que la auditoría interna sea útil en la implementación juiciosa de políticas y procedimientos administrativos.

Con toda seguridad, a medida que la cultura organizacional mejora y la calidad de la autoevaluación aumenta, la evaluación independiente debería apartarse de los dominios que la administración maneja mejor y reorientar su foco; por ejemplo, puede moverse a una posición superior que la autoevaluación todavía no atiende adecuadamente. En otras palabras, la «subsidiariedad» —el principio de organización por el cual los asuntos deberían manejarse en el nivel de autoridad competente más bajo y menos centralizado— es un excelente principio para el diseño de la relación entre evaluación independiente y autoevaluación.

La autoevaluación y la evaluación independiente deberían estar muy conectadas

Una combinación de evaluación independiente y de autoevaluación estimula a los administradores a diseñar programas evaluables, es decir, la adopción de metas claras, objetivos verificables y planes de monitoreo y de evaluación adecuadamente financiados. A modo de ejemplo, los enfoques basados en gestión por resultados, para que la evaluación funcione, debería incorporarse por adelantado en el diseño de las intervenciones. Esto debería ser parte integrante de la garantía de calidad para todas las intervenciones públicas.

La provisión regular de información validada independientemente sobre cómo la organización usa sus recursos y si sus intervenciones abordan prioridades clave y si obtienen resultados o no, implica que debería realizarse un seguimiento a todos los principales programas y que la información precisa de la evaluación debería publicarse con regularidad, de manera que las partes interesadas tengan acceso a información clara, regular y creíble sobre la eficacia organizacional.

En el mejor de los casos, el monitoreo está orientado a la evaluación y reevaluación de políticas y programas. Esto significa que el monitoreo

es crítico para la administración, pero también que tanto la evaluación independiente como la autoevaluación agregan valor al monitoreo aplicando criterios y estándares a los hallazgos empíricos, examinando el mérito y el valor de los resultados, explicando las desviaciones de los planes, teniendo en cuenta los cambios en el contexto, reconsiderando los supuestos en los que se basó la intervención original, comprobando la justificación continua de la intervención, guiando la toma de decisiones con respecto a su adaptación a nuevas circunstancias, enfrentando las consecuencias no buscadas, realineando los roles y responsabilidades de los asociados, entre otros.

¿El tamaño importa?

En la evaluación, menos puede ser más. Se establecen rendimientos decrecientes con rapidez cuando la evaluación independiente aumenta los costos de transacción sin inducir mucho la rendición de cuentas o el aprendizaje. En términos de Kenneth Arrow, «Para cumplir con sus funciones, la responsabilidad debe ser capaz de corregir los errores, pero no debe ser tal que destruya los valores genuinos de la autoridad. Está claro que un órgano de responsabilidad lo suficientemente estricto y continuo puede equivaler con facilidad a una negación de la autoridad (...). Para conservar el valor de la autoridad, la responsabilidad debería ser intermitente» (Arrow, 1974: 77-78).

Por consiguiente, la autoevaluación que está al servicio y fortalece la autoridad, y reduce el índice de innecesarios errores organizacionales debería privilegiarse en la asignación de recursos. Asimismo, la función de evaluación que combina evaluación independiente y autoevaluación no debería ser privada de recursos. Y deberían asignarse presupuestos suficientes que permitan la valoración justa, válida, precisa y bien documentada de toda la eficacia organizacional.

Con este propósito, los sistemas de monitoreo y de autoevaluación deberían complementarse con evaluaciones independientes que realicen controles aleatorios y atestigüen la validez de las afirmaciones de la autoevaluación. Una masa crítica de recursos asignados a ambas funciones es necesaria para que pueda responsabilizarse a la autoridad. Sin presupuestos adecuados, la evaluación corre el riesgo de quedar relegada a un papel simbólico.

Conclusiones

La evaluación independiente ayuda a mejorar el desempeño organizacional. En última instancia, esto es así porque el aprendizaje organizacional difiere del aprendizaje individual. En particular, en las organizaciones grandes y jerárquicas en las que las lecciones extraídas no son necesariamente lecciones aprendidas. Los sólidos hallazgos de la evaluación no se adoptan automáticamente. Más allá de la adquisición de conocimiento, el proceso de aprendizaje requiere cambios actitudinales y del comportamiento. Para las organizaciones, esto implica cambios periódicos en la estrategia, las políticas y los procedimientos.

Se requiere inducir al personal a observar los procedimientos acordados y a enfocarse en las metas empresariales para alcanzar la eficacia organizacional: este es el terreno del monitoreo, la auditoría y el aprendizaje unidireccional. Pero el aprendizaje bidireccional también es necesario, es decir, las estrategias, políticas, procesos y procedimientos que moldean el comportamiento organizacional requieren un ajuste oportuno cuando el entorno operativo está turbulento o inestable: esto se facilita con la ayuda de las evaluaciones independientes de alta calidad que interrogan la validez continua de los objetivos estratégicos establecidos y al reconsiderar el fundamento de los procesos de negocios profundamente arraigados.

En cambio, los sistemas de gestión que no contribuyen con la adaptación oportuna de las metas estratégicas ni con los lineamientos operativos son riesgosos. Hacer las cosas bien puede destruir a las organizaciones que no hacen lo correcto. La evaluación integrada a los procesos operativos ayuda a asegurar los ajustes ágiles de estrategias y políticas a los cambios en el entorno externo.

La autoevaluación está más cerca de la acción y es más probable que impacte en el comportamiento organizacional. No obstante, carece de la distancia necesaria para una valoración objetiva, por lo que es posible que su calidad se deteriore, y es vulnerable a la captación de los intereses creados en la organización sin una función de evaluación independiente que se encargue de atestiguar la validez de sus hallazgos, como la auditoría respecto a la contabilidad.

En resumen, la independencia de evaluación definida en términos funcionales y no simplemente de las habilidades y de la disposición de los evaluadores es un ingrediente esencial de la excelencia de la evaluación. Ayuda a mejorar la calidad y la credibilidad de los productos de la evaluación y contribuye a la transparencia y la rendición de cuentas organizacional. El concepto se codificó para asegurar que los evaluadores estén equipados para brindar informes intransigentes y de alta calidad, y para proteger la función de la captación y de la intimidación. Todos los criterios de la independencia de evaluación (protección organizacional y comportamental, protección de influencias externas y elusión del conflicto de interés) deben ponerse en marcha para que la función sea genuinamente independiente.

Sin embargo, independencia no es aislamiento. De hecho, la evaluación no facilita el aprendizaje organizacional si no logra contribuir con la formulación estratégica, amplificar la voz de las partes interesadas legítimas o proporcionar información creíble y confiable del desempeño a la máxima autoridad de gobernanza. La evaluación independiente también necesita apalancar su impacto por medio de vínculos apropiados con los procesos de autoevaluación con el fin de generar «aprendizaje bidireccional».

Esto implica conexiones compatibles entre evaluación independiente y autoevaluación, incluidos los protocolos juiciosos de la interacción profesional. También, un cuidadoso diseño institucional de las estructuras organizacionales y de los procesos comerciales, ya que sin relaciones eficaces entre las unidades de supervisión empresarial a cargo del aprendizaje unidireccional y las unidades de evaluación encargadas del aprendizaje bidireccional puede haber incoherencia y duplicación. Por último, no hace falta decir que tanto la evaluación independiente como la autoevaluación deberían ser financiadas adecuadamente para tener un impacto significativo en la organización y para evitar la percepción de que solo son una fachada.

Notas

* Título original "Evaluation Independence in organizations", en *Journal of MultiDisciplinary Evaluation*, Volumen 9, N.º 20, 2013, ISSN 1556-8180. Publicación autorizada en inglés ©2013 *Journal of MultiDisciplinary Evaluation*. Traducción autorizada al español. Todos los derechos reservados.

Reconocimientos: El presente artículo se basa en un ensayo que encargó y respaldó la Oficina de Evaluación Independiente del Fondo Monetario Internacional en 2012. El autor le agradece a Matt Galen por sus útiles comentarios y por su generosa colaboración editorial.

1. El uso de los enfoques basados en la evaluación destacan el poder de los procesos de evaluación colaborativos y participativos que construye confianza y obtiene la atención, conocimiento y compromiso de los gerentes y personal del programa (Patton, 1997).

2. El «Glosario de los términos clave sobre evaluación y gestión basada en resultados» confeccionado por el Comité de Asistencia al Desarrollo de la OCDE (Organización para la Cooperación y el Desarrollo Económicos) especifica que una evaluación es independiente cuando «es realizada por entidades y personas que no están sometidas al control de aquellos que son responsables del diseño y de la implementación de la intervención del desarrollo». También indica que la evaluación independiente presume «libertad de influencia política y de presión organizacional», «total acceso a la información» y «completa autonomía para llevar adelante investigaciones y reporte de los hallazgos».

Referencias

Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review*, 55(5), 115-125.

Arrow, K. J. (1974). *The limits of organization*. Nueva York, Nueva York: Norton.

Evaluation Cooperation Group (2010). Good practice standards on independence of international financial institutions' central evaluation departments. Obtenido de [https://wpqr4.adb.org/lotusquickr/ecg/Main.nsf/0/BEB596AD98892E95482579C70023C0C1/\\$file/Public%20Sector%20GPS%2004-19-12%20_final.pdf](https://wpqr4.adb.org/lotusquickr/ecg/Main.nsf/0/BEB596AD98892E95482579C70023C0C1/$file/Public%20Sector%20GPS%2004-19-12%20_final.pdf).

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to declines in firms, organizations, and states*. Cambridge, Massachusetts: Harvard University Press.

Love, A. J. (1991). *Internal evaluation: Building organizations from within*. Newbury Park, California: SAGE.

- Mayne, J. (2008). *Building an evaluative culture for effective evaluation and results management*. (Documento de trabajo N.º 8). Roma, Italia: Institutional Learning and Change Initiative. Obtenido de http://www.focusintl.com/RBM107-ILAC_WorkingPaper_No8_EvaluativeCulture_Mayne.pdf
- Moore, M. H. (1995). *Creating public value: Strategic management in government*. Cambridge, Massachusetts: Harvard University Press.
- Olson, M., Jr. (1971). *The logic of collective action: Public goods and the theory of groups*. Cambridge, Massachusetts: Harvard University Press.
- Perkins, D. N. (1993). Teaching for understanding. *American Educator*, 17(3), 28-35.
- Picciotto, R. (2005). The value of evaluation standards: A comparative assessment. *Journal of Multidisciplinary Evaluation*, 2(3), 30-59.
- . (2011). The logic of evaluation professionalism. *Evaluation*, 17(2), 165-180.
- Pollitt, C. y Boukaert, G. (2000). *Public management reform: A comparative analysis*. Oxford, Inglaterra: Harvard University Press.
- Scriven, M. (1991). *Evaluation thesaurus*. Cuarta Edición. Newbury Park, California: SAGE.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick, Nueva Jersey: Transaction.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21-33.

Anexo

Plantilla para valorar las organizaciones con independencia de evaluación

| criterio | Aspectos | Indicadores |
|--|---|--|
| I. Independencia organizacional | La estructura y el papel de la unidad de evaluación. | Si la unidad de evaluación tiene una declaración de mandato que aclara que el alcance de su responsabilidad abarca todas las operaciones de la organización y que su línea de reporte, personal, presupuesto y funciones son organizacionalmente independientes de los departamentos operativos, de políticas y de estrategia de la organización, así como de la toma de decisiones relacionada. |
| | La unidad rinde cuentas y presenta los resultados de la evaluación al director o subdirector de la organización o a su junta directiva. | Si hay una relación directa con respecto al reporte entre la unidad y: <ul style="list-style-type: none"> • la administración. • la junta. • el comité correspondiente de la junta de la institución. |
| | La unidad se encuentra organizacionalmente fuera de la función de gestión del personal o de la línea del programa, la actividad o la entidad que se evalúa. | La posición de la unidad en la organización con respecto al programa, actividad o entidad que se evalúa. |
| | La unidad reporta regularmente al comité de auditoría o a otro órgano de supervisión de la organización. | Relación y frecuencia de reporte ante el órgano de control. |
| | La unidad está bastante alejada de las presiones políticas para poder informar los hallazgos sin temor a las repercusiones. | Medida en la que la unidad de evaluación y su personal deben rendir cuentas a las autoridades políticas y en que están protegidos de la participación en actividades políticas. |
| | Los integrantes de la unidad están protegidos por un sistema de personal en el que remuneración, capacitación, cargos y ascensos se basan en el mérito. | Medida en la que existe y se impone un sistema de mérito que cubre remuneración, capacitación, cargos y ascensos. |
| | La unidad tiene acceso a toda la información necesaria y a las fuentes de información. | Medida en la que la unidad de evaluación tiene acceso a: <ul style="list-style-type: none"> • personal, registros y emplazamiento de proyectos. • cofinancistas y otros asociados, clientes. • programas, actividades o entidades que se financian o patrocinan. |
| II. Independencia de comportamiento | Capacidad y disposición para elaborar informes sólidos, de alta calidad e intransigentes. | Medida en la que la unidad de evaluación: <ul style="list-style-type: none"> • elabora informes de alta calidad que inviten al escrutinio público (dentro de salvaguardias adecuadas para proteger información confidencial y de propiedad, y para mitigar riesgo institucional) de las lecciones de los programas y actividades de la organización. • propone estándares de desempeño más adelantados que los que la organización usa en la actualidad. • critica resultados de los programas, actividades y entidades de la organización. |

Plantilla para valorar las organizaciones con independencia de evaluación (cont.)

| Criterio | Aspectos | Indicadores |
|--|--|---|
| | Capacidad para informar con franqueza. | Medida en la que el mandato de la organización establece que la unidad de evaluación transmite sus informes a la administración o junta después de revisión y comentarios de unidades empresariales relevantes, pero sin restricciones impuestas por administración sobre su alcance y comentarios. |
| | Transparencia en el reporte de los hallazgos de la evaluación. | Medida en la que las reglas de divulgación de la organización permiten que la unidad de evaluación informe hallazgos significativos a partes interesadas afectadas, tanto internas como externas (dentro de las salvaguardias adecuadas para proteger información confidencial o de propiedad, y para mitigar riesgo institucional). |
| | Autoselección de temas para programa de trabajo. | Procedimientos para selección de temas del programa de trabajo son elegidos, por medios sistemáticos o intencionales, por la organización de evaluación; se consulta con la Administración y con la Junta sobre el programa de trabajo. Quién determina la política y procedimientos de divulgación de la unidad de evaluación: la junta, el comité correspondiente o la administración. |
| | Protección del presupuesto administrativo y de otras fuentes presupuestarias para la función de evaluación. | Partida del presupuesto administrativo para evaluación determinada de acuerdo con parámetro claro en materia de políticas y preservada a un nivel o proporción indicado; acceso a fuentes adicionales de financiamiento solo con revisión formal del contenido de las presentaciones. |
| III. Protección de influencia externa | Diseño y ejecución adecuados de la evaluación. | Medida en la que la unidad de evaluación puede determinar el diseño, alcance, ritmo y conducción de las evaluaciones sin influencia de la Administración. |
| | Financiación del estudio de la evaluación. | Medida en la que la unidad de evaluación no está impedida por restricciones presupuestarias u otros recursos que afectarían negativamente su capacidad para realizar sus responsabilidades. |
| | Juicios hechos por los evaluadores. | Medida en la que el juicio del evaluador sobre el contenido adecuado de un informe no está sujeta a revocación o influencia de autoridad externa. |
| | Contratación y despido del director de la unidad de evaluación, duración del mandato, revisión del desempeño y remuneración. | El mandato o documento equivalente especifica procedimientos para: <ul style="list-style-type: none"> • contratación y despido. • duración del cargo. • revisión del desempeño. • remuneración del director de la unidad de evaluación que asegura su independencia de la administración operativa. |
| | Contratación, promoción o despido del personal. | Medida en la que la unidad de evaluación tiene control sobre: <ul style="list-style-type: none"> • contratación del personal. • promoción y aumentos de salario. • despido, dentro de un sistema de mérito. |

Plantilla para valorar las organizaciones con independencia de evaluación (cont.)

| Criterio | Aspectos | Indicadores |
|---|--|---|
| | Empleo continuo del personal. | Medida en la que la continuidad del empleo del evaluador se basa solo en motivos relacionados con desempeño laboral, competencia o necesidad de los servicios del evaluador. |
| IV. Elusión de conflictos de interés | Relaciones oficiales, profesionales, personales o financieras que podrían causar que un evaluador limite el alcance de una investigación, acote la divulgación o debilite o sesgue los hallazgos. | Medida en la que hay políticas y procedimientos establecidos identifican las relaciones del evaluador que podrían interferir con la independencia de la evaluación; se comunican al personal mediante capacitación y otros medios; además, son impuestos. |
| | Ideas preconcebidas, prejuicios o sesgos sociales o políticos que podrían afectar los hallazgos de la evaluación. | Medida en la que las políticas y procedimientos están establecidos y son impuestos que requieren de los evaluadores: <ul style="list-style-type: none"> • valorar e informar prejuicios o sesgos personales que hacen peligrar su capacidad para aportar objetividad a la evaluación. • determinar qué partes interesadas son consultadas como parte del proceso de evaluación para asegurarse contra los sesgos del evaluador. |
| | Participación actual o previa en un programa, actividad o entidad que se evalúa a un nivel de toma de decisiones o en una función de administración financiera o contable, o la búsqueda de empleo en ese programa, actividad o entidad mientras se conduce la evaluación. | Medida en la que reglas o procedimientos para el personal se presentan o se imponen para evitar que este evalúe programas, actividades o entidades para los que tiene o tuvo un papel vinculado con toma de decisiones o administración financiera, o en los cuales busca empleo. |
| | Interés financiero en el programa, actividad o entidad que se evalúa. | Medida en la que reglas o procedimientos para el personal se establecen y se imponen para evitar que el personal evalúe programas, actividades o entidades en los que sus familiares tienen interés financiero. |
| | Familiar inmediato o cercano está involucrado o en posición para ejercer influencia directa y significativa en el programa, actividad o entidad que se evalúa. | Medida en la que reglas o procedimientos para personal se establecen y se imponen para evitar que el personal evalúe programas, actividades o entidades en los que sus familiares ejercen influencia. |

Nota: adaptación de «Estándares de Auditoría de Gobierno, Enmienda 3», de la Oficina de Rendición de Cuentas del Gobierno de los Estados Unidos, 2002; «Glosario de los principales términos sobre evaluación y gestión basada en resultados», de la Organización para la Cooperación y el Desarrollo Económicos, 2002; «Principios de Evaluación de la Asistencia al Desarrollo», de la Organización para la Cooperación y el Desarrollo Económicos, 1991; «Código de Ética y Estándares de Auditoría», de la Organización Internacional de las Entidades Fiscalizadoras Superiores, 2001; «Marco de Prácticas Profesionales», del Instituto de Auditores Internos, 2000; «El enfoque conceptual para proteger la independencia del auditor», de la Federación Europea de Contadores, 2001; «Lineamientos de Evaluación», del Ministerio de Relaciones Exteriores de Dinamarca, 1999; «Guía de Evaluación», de la Agencia Canadiense para el Desarrollo Internacional, 2000.

La Serie *Estado, Gestión Pública y Desarrollo en América Latina* presenta una selección de textos dirigida a fortalecer la comprensión e implicancias que el funcionamiento del Estado y de las instituciones en América Latina tiene para el desarrollo.

Con este objetivo, los textos brindan información sobre los fundamentos del debate teórico internacional con foco en estas cuestiones, así como también transmiten análisis de experiencias significativas de nuestras sociedades y de otras latitudes con miras a difundir estos conocimientos en nuestra región.

CAF, como institución dedicada a promover el desarrollo, pone esta colección a disposición de quienes participan activamente de la gestión pública, así como de académicos, estudiantes y de todos aquellos con interés en la materia, con el propósito de acercarlos esta literatura y apoyarlos en el proceso de comprensión, diseño y ejecución de políticas públicas asociadas a la calidad institucional y al desarrollo en nuestros países.

La evaluación de políticas. Fundamentos conceptuales y analíticos

La evaluación de políticas se ha convertido en un campo estratégico para el mejoramiento de las políticas públicas en América Latina, con gran vitalidad e intensos debates disciplinarios, teóricos y metodológicos. No obstante estos avances, la comprensión de los desafíos que esta práctica enfrenta no se manifiesta de manera semejante en los diferentes países, ámbitos sectoriales y organizaciones de gobierno. El escenario actual aún requiere la consolidación de su institucionalización y una mayor utilización de sus resultados en el diseño y la implementación de políticas y programas públicos.

La evaluación plantea desafíos muy específicos, destacados en este libro. Algunos son propios de la naturaleza política de los múltiples juegos en que intervienen los actores sociales en torno a las políticas públicas; otros corresponden a la problemática del conocimiento que es movilizado a lo largo de los procesos de definición de problemas, de generación y consideración de alternativas, de formalización y de gestión de las políticas y de identificación y apreciación de sus consecuencias.

Es en este contexto que el volumen pone a disposición del lector temas y enfoques referidos a la evaluación de políticas y programas públicos. Su propósito es presentar y discutir fundamentos epistemológicos, conceptuales y analíticos como un paso ineludible para el enriquecimiento de las intervenciones dirigidas a la institucionalización de la evaluación en América Latina.

Datos de contacto de la Dirección Corporativa, Desarrollo Institucional: desarrolloinstitucional@caf.com